AFRL-RH-WP-TO -2010-2226

Compexity in Glyphs and Systems
(Collection of Papers)

Daniel W. Repperger

Human Effectiveness Directorate
711th Human Performance Wing
Wright-Patterson AFB OH 45433

APRIL 2010
Final Report

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

//signed//
DOUGLAS FRANCK
Technical Advisor
Battlespace Visualization Branch

//signed//
JEFFREY L. CRAIG
Chief, Battlespace Visualization Branch
Warfighter Interface Division

//signed//
MICHAEL A. STROPKI
Chief, Warfighter Interfaces Division
Human Effectiveness Directorate
711th Human Performance Wing

REPORT DOCUMENTATION PAGE		*Form Approved* OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 04-05-2010	2. REPORT TYPE Final	3. DATES COVERED *(From - To)* 2 Jun 2005 – 5 Apr 2010

4. TITLE AND SUBTITLE Complexity in Glyphs and Systems	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBERS 61102F

6. AUTHOR(S) Repperger, D. W.	5d. PROJECT NUMBER 2313
	5e. TASK NUMBER HC
	5f. WORK UNIT NUMBER 2313HC54

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) 711 HPW/RHCV 2255 H Street Wright-Patterson AFB OH 45433	8.PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711th Human Performance Wing Human Effectiveness Directorate Warfighter Interface Division Battlespace Visualization Branch Wright-Patterson AFB OH 45433-7022	10.SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHCV
	11.SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-RH-WP-TM-2010-0004

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Collection of papers: Paper 1: Modifying Sensitivity/Specificity for Sensors Using Positive and Negative Predictive Power Measures (88 ABW Cleared 06/17/2009; 88ABW-09-2618), Paper 2: Updated Studies in Approximate Entropy involving Fractional Noise and Fatigue Data, and Paper (88 ABW Cleared 09/16/2009; 88ABW-09-4068), 3: On the Absolute Orientation Problem in Computer Vision (88 ABW Cleared 06/17/2009; 88ABW-09-2613)

14. ABSTRACT
To advance basic research in display and systems that is required by improving the mathematical underpinnings. For visual display icon renderings, how complex should they be synthesized to effectively provide information to the decision maker. For complex military network systems, how can we identify vulnerability and performance in a quantitative manner. In addition, can objectively quantify the information overload limits of humans.

15. SUBJECT TERMS warfighter-system interface, battlespace visualization, approximate entropy, glyphs, complex systems, man-machine interfaces, mathematical structure, uncertainty, visualizations, and computer vision

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGE	19a. NAME OF RESPONSIBLE PERSON Douglas L. Franck
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR	37	19b. TELEPONE NUMBER *(Include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

This page intentionally left blank.

Modifying Sensitivity/Specificity for Sensors Using Positive and Negative Predictive Power Measures

D. W. Repperger, J. S. Warm, P. R. Havig, K. Farris
711 HPW AFRL WPAFB
Dayton, Ohio USA

M. A. Vidulich, V. S. Finomore
711 HPW AFRL WPAFB
Dayton, Ohio USA

Abstract — **In the collection of data from sensors in the field, the u ncertainty in th e d ata may comp romise th e ab ility to accurately p redict th e s tate of a s ystem. H erein th e s tandard signal detection theory problem is examined when nonstationary effects may occur in the data from the sensors. The use of *PPP* (positive predictive power) and *NPP* (negative predictive power) adds a ne w view point on how   to modify   se nsitivity an d specificity mea sures in d ecision making involving multiple sensors. Th is is es pecially t rue when s tationary p roperties in received data may be violated.**

I.   INTRODUCTION

In standard signal detection theory (SDT) analysis, the presumption is usually that the process acts in a stationary manner. The noise processes which permeate the data may be presumed to be Gaussian with a constant mean and variance. In real applications, however, the actual uncertainty modeled via randomness may act more like a stochastic process with a time varying mean and standard deviation. These variations in the first and second statistical moments commonly occur, e.g. when sensors heat up, cool down, or have other time variations. The sensors may even fail. If the quality of data are changing from the sensors, one must modify the decision making process to account for uncertainty induced by a compromised time varying data stream. This paper will examine decision making when the quality of data are adapting with time and use measures such as *PPP* or *NPP* to improve upon current decision-making methodologies based on static assumptions.

II.   BACKGROUND MATERIAL

(2.A) Definition of a Stochastic Process

Figure (1) portrays the concept of a stochastic process [1]. The x axis is the independent variable, time. The vertical axis is for the dependent variable of the stochastic process

y(t). This stochastic process may have a first moment (mean $\mu(t)$) which changes with time. The square root of the second

moment (standard deviation $\sigma(t)$) also may vary with time. These variations are portrayed in Figure (1) on the vertical axis in the form of probability density functions which changes temporally.

Figure (1) – A Rendering of a Stochastic Process

The basics of signal detection theory will be discussed next.

(2.B) – A Signal Detection Theory Framework

Figure (2) – Confusion Matrix for Statistical Test

Figure (2) portrays a confusion matrix for a binary detection task. In Figure (2), let H denote the number of hits, CR is correct rejects, FA denotes false alarms and M represents the number of misses. The standard definitions that will be used here are described in equations (1- 4) [2]:

$$\text{Sensitivity} = Sn = \frac{H}{H+M} \tag{1}$$

$$\text{Specificity} = Sp = \frac{CR}{CR+FA} \tag{2}$$

$$\text{Positive Predictive Power} = PPP = \frac{H}{H+FA} \tag{3}$$

$$\text{Negative Predictive Power} = NPP = \frac{CR}{CR+M} \tag{4}$$

(Note $Sn \neq PPP$ and $Sp \neq NPP$ from equations (1-4).)

(2.C) The Use of PPP and NPP to Assist in Decision Making

Figure (3) shows the typical statistical test for a binary detection task and the appropriate parameters to discern
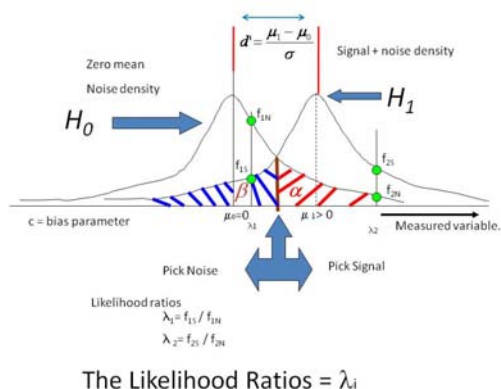

Figure (3) – Classical Binary Decision Making with Optimality

objects with the signal to noise ratio d' and bias parameter c as noted. In the medical diagnostics area physicians now highly embrace the concepts of *PPP* and *NPP* rather than just using d' and c for several important reasons: (1) The data properties may change with time with d' and c adapting [3,4]. Also, (2) d' and c are variables calculated with a static analysis, however, the underlying system may experience dynamic changes requiring modification of the decision making process. Evidence of nonstationary behavior in decision making is pervasive in the literature on experimental psychology.

(2.D) – Extant Data in Human Decision Making.

In the literature, there is extensive evidence indicating that human decision making is a time-varying process. In studies on fatigue or vigilance, this effect is seen quite often. For example [3-5] it is shown that d' in Figure (3) may vary with

time, but *PPP* and *NPP* may remain constant. Also in [6, 7] evidence suggests that dynamic changes occur in decision making, especially in studies in vigilance and fatigue. In [4] it was shown that *PPP* may be constant (performance feedback provided) but *NPP* may decrease with time. Thus *PPP* more specifically delineates **what types of decisions** have higher credence. This is a stronger result than what is provided by the typical static measures from signal detection theory and has achieved high acceptance in the medical community [10].

III.   SOME KEY CLASSICAL RELATIONSHIPS

To relate the terms in equations (1-4) to some of the well known quantities in statistical testing for binary hypotheses, the elements of figure (2) can be described via [8]:

$$\alpha = \frac{FA}{CR+FA} = 1 - specificity \tag{5}$$

or
$$specificity = 1 - \alpha \tag{6}$$

and
$$\beta = \frac{M}{M+H} = 1 - sensitivity \quad = \text{power of the test} \tag{7}$$

with
$$sensitivity = 1 - \beta \tag{8}$$

Note, also in Figure (3):
$$d' = \frac{\mu_1 - \mu_0}{\sigma} \tag{9}$$

(3.A)   Optimality Measures for decision making

Three types of optimality in decision making are briefly discussed herein. The first two types of decision making can be developed using the d' and c values in Figure (3).

Optimality Test -1 – Neyman-Pearson (fix α, minimize β)

From Figure (3) and the classical maximum likelihood ratio test, it is well known [2] that using a decision rule based on the likelihood ratio is optimal in the Neyman-Pearson sense. This means for a fixed α (type 1 error), the minimum type 2 error (β in Figure (3)) is realized. The decision rule employs the ratio of the probability density functions in Figure (3). The decision rule is to select choice H_1 over choice H_0 if (threshold = unity in Figure (3)):

$$\lambda = (f_1/f_0) > \text{Threshold} = 1 \tag{10}$$

The second and well accepted test of optimality involves the area under the ROC (receiver operator characteristic) curve.

Optimality Test 2 – Area under an ROC curve

In the fields of medical decision making and diagnostisity in general, a concept called "discriminability" is highly touted in these fields. Herein a brief description of the ROC curve is provided to be instructive. Figure (4) shows the ROC curve as
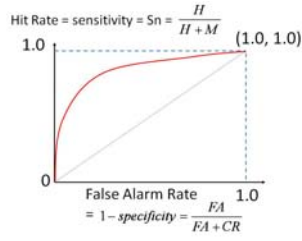
2

Figure (4) - The ROC (Receiver Operator Characteristic) Curve

a plot of sensitivity versus 1-specificity. The area under the ROC curve (discriminability) accounts for the general ability of a statistical test to include the two types of errors (misses and false alarms). It represents an overall measure of the efficacy of the test. This is precisely what *PPP* and *NPP* can bring to the decision making process.

From Figure (4), the efficacy of the test (discriminability) is the total area under the ROC curve which is desired to be maximized. It should be pointed out that ROC curves are not arbitrary functions and have certain restrictive properties.

The next optimality test will generalize signal detection theory to dynamic systems which is appropriate if the moments μ(t) and σ(t) have time variations.

Optimality Test 3 - The Wald Sequential Test

The classical Wald sequential algorithm test involves two types of optimality and provides an entry into dynamic decision making [9]. Figure (5) portrays a dynamic decision making situation in which the ROC curve may be changing with time.
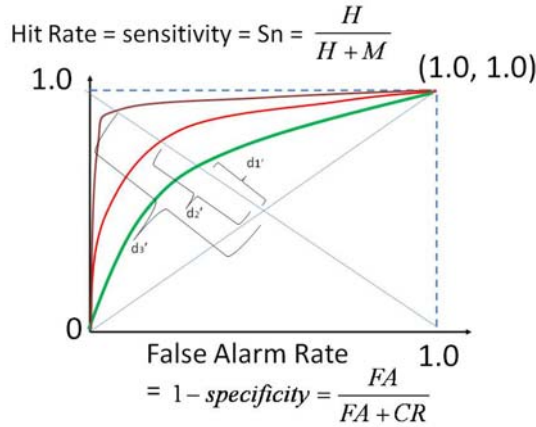

Figure (5) – Adapting ROC curves as System Parameters Vary

In Figure (5) the d' and c parameters may vary with time and the goal is to adjust the decision rule, accordingly. To complete this background information, the optimal Wald sequential decision test can be briefly described as follows [9]:

A statistical measure $\gamma(t)$ (likelihood ratio) is computed in real time (Wald test), where

$$A(t) = \log\frac{\beta}{1-\alpha} \ \leq \ \gamma(t) \ \leq \ \log\frac{1-\beta}{\alpha} = B(t) \quad (11)$$

The decision rule is: If the determined $\gamma(t) > B(t)$, then select hypothesis H_1 in Figure (3). However, if $\gamma(t) < A(t)$, then choose H_0. If neither these conditions are true, then collect more data.

The beauty of the Wald test is that if α or *β* vary with time, then the statistical parameters A(t) and B(t) adjust accordingly. The decision making process then modifies. It is shown [9] that the Wald test still maintains two types of optimality: (1) It is still Neyman-Pearson optimal in the sense that for a fixed α (type 1 error) then β (type 2 error) is minimized, as well as (2) the algorithm will converge in minimum time to a decision. What this means is that a decision is reached in about 50% of the time [9] as compared to a static maximum likelihood decision making process with constant values of *α* and *β*. Thus dynamic decision making with time varying variables has significant advantages over static methodologies.

Finally, the concepts of *PPP* and *NPP* can now be generalized into dynamic decision making employing the background materials developed so far.

IV.   ADAPTING DECISION MAKING WITH *PPP* AND *NPP*

Additional facts on the derivation and relationships of *PPP* and *NPP* to dynamical statistical hypothesis testing are detailed here. A brief description of the pertinent steps will be presented here. Equations (12-19) now show key relationships between the classical quantities *CR, M, H,* and *FA* from Figure (1) to the *PPP* and *NPP* parameters as well as sensitivity (*Sn*) and specificity (*Sp*).

$$CR = M\left(\frac{NPP}{1-NPP}\right) \quad (12)$$

$$M = CR\left(\frac{1-NPP}{NPP}\right) \quad (13)$$

$$H = (FA)\left(\frac{PPP}{1-PPP}\right) \quad (14)$$

$$FA = (H)\left(\frac{1-PPP}{PPP}\right) \quad (15)$$

$$CR = (FA)\frac{Sp}{1-Sp} \quad (16)$$

$$FA = (CR)\frac{1-Sp}{Sp} \quad (17)$$

$$H = (M)\frac{Sn}{1-Sn} \quad (18)$$

$$M = (H)\frac{1-Sn}{Sn} \quad (19)$$

3

The following key result has its details contained in Appendix A and can be described via Theorem 1:

Theorem 1: Sensitivity (*Sn)* and specificity (*Sp*) can be related to *PPP* and *NPP* through the following formula:

$$\left(\frac{Sn}{1-Sn}\right)\left(\frac{Sp}{1-Sp}\right) = \left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right) \qquad (20)$$

Proof: Appendix A derives the result in equation (20).

The second major result is to extend the Wald sequential test from equation (11) into a form for the use of *PPP* and *NPP*. With this dynamic decision making criterion the Neyman-Pearson property of optimality is still preserved and the minimum time for a decision to be made is still left intact. Theorem 2 presents this main result:

Theorem 2:

The equivalent to the Wald sequential test of equation (11) can be written in terms of *PPP* and *NPP* via the following method:

The optimal Wald sequential algorithm can be modified to include the *PPP* and *NPP* terms. For the test statistic *γ(t)* (likelihood ratio) in equation (11), the new regions for testing now include:

$$\log\left[\left(\frac{1-\beta}{\alpha}\right)\left(\frac{1-PPP}{PPP}\right)\left(\frac{1-NPP}{NPP}\right)\right] \le \gamma(t) \le$$
$$\log\left[\left(\frac{\beta}{1-\alpha}\right)\left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right)\right] \qquad (21)$$

Proof: Appendix B outlines the derivation of equation (21). Equation (21) is equivalent to equation (11) and still preserves the two types of optimality as enjoyed by the Wald sequential test.

Remarks: The formulation (21) has new advantages over prior methods. It is known that the parameters d' and c of Figure (2) may vary with time. From equation (21) it is seen that the decision parameters *α* and *β* are known *apriori*. Also the experimental parameters *PPP* and *NPP* adapt with time since the d' and c values are known to change but are easily measured in real time. Thus by using equation (21) rather than (11) provides a dynamic decision making procedure which can be modified with time having an emphasis on the time varying *PPP* and *NPP* variables. The result in equation (21) still enjoys the same optimality properties as in the Wald sequential algorithm in equation (11) but is cast within the framework of the *PPP* and *NPP* variables.

V. SUMMARY AND CONCLUSIONS

Sensitivity and specificity using a *PPP* and *NPP* formulation demonstrate that a dynamic Wald-type sequential algorithm can be synthesized depending only on the *PPP, NPP*, *α* and *β* values. From equation (20), several new viewpoints on dynamic decision making can be obtained. For example, if *PPP* may be constant and *NPP* may be decreasing (cf. [4]), then both sensitivity and specificity may change.

This because from equations (1,3) *PPP* and *Sn* are different quantities. The new viewpoint allows us to drill down on the decision making when overall measures like (d', c, *Sn* and *Sp* may be changing with time. Also, the new dynamic decision making algorithm presented herein preserves the two types of statistical optimality provided by the Wald method but uses more modern procedures such as *PPP* and *NPP*.

REFERENCES

[1] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th Edition, McGraw Hill, 2002.
[2] Swets, J.A. *Signal detection theory and ROC analysis in psychology and diagnostics*: Collected papers. Lawrence Erlbaum Associates, 1995.
[3] Warm, J. S. et al., "Effects of Training with Knowledge of Results on Diagnosticity in Vigilance Performance," Human Factors and Ergonomics Society 53rd Annual Meeting, San Antonio, Texas, October, pp. 192-23, 2009.
[4] W. N. Dember and J. S. Warm, *Psychoogy of Perception*, Second Edition, Holt, Rinehart and Winston, 1979.
[5] M. J. Hausen, "Training for Vigilance: Effects on Performance Diagnosticity, Stress, and Coping," MA Thesis, Department of Psychology, University of Cincinnati, 2008.
[6] J. S. Warm, R. Parasuraman, and G. Matthews, "Vigilance Requires hard Mental Work and is Stressful," Human *Factors*, **50** (3), June, 2008, pp. 433-441.
[7] W. S. Helton, G. Matthews, and J. S. Warm, "Stress state mediation between environmental variables and performance: The case of noise and vigilance," *Acta Psychologica*, **130**(2008), pp.204-213.
[8] D. W. Repperger, D. L. Aleva, G. Thomas, J. E. Miller, and S. C. Fullenkamp, "Complexity of Visual Icons Studied via Signal Detection Theory," *Perceptual and Motor Skills – Psychological Reports*, **vol. 105**, pp. 287-298, October, 2007.
[9] Wald, Abraham, "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics,* **16** (2), 1945, 117–186.
[10] S. Linton, "General Internal Medicine,: in U. B. S. Prakash (Ed.), *Mayo Internal Medicine Board Review* (pp. 333-352). Rochester, MN: Mayo Clinic, 1996.

Appendix A – Derivation of Equation (20)

Starting with the equation (4), multiplying through by *(CR+M)* yields:

$$(CR+M)\, NPP = CR \qquad (A.1)$$
or $\qquad CR\,(1\text{-}NPP) = M\,(NPP) \qquad (A.2)$
which gives rise to the following two relationships:

$$CR = M\left(\frac{NPP}{1-NPP}\right) \qquad (A.3)$$

$$M = CR\left(\frac{1-NPP}{NPP}\right) \qquad (A.4)$$

Starting with equation (3), multiply through by (*H+FA*) to yield:

$$(H+FA)\, PPP = H \qquad (A.5)$$
or $\qquad H\,(1\text{-}PPP) = FA\,(PPP) \qquad (A.6)$

4

which arrives at the following two relationships:

$$H = (FA)\left(\frac{PPP}{1-PPP}\right) \tag{A.7}$$

$$FA = (H)\left(\frac{1-PPP}{PPP}\right) \tag{A.8}$$

To continue, start with equation (2) and multiply through by $(CR+FA)$ to yield:

$$Sp\,(CR + FA) \;=\; CR \tag{A.9}$$

or: $\qquad CR\,(\,1 - Sp\,) = Sp\,FA \tag{A.10}$

which results in the following relationships:

$$CR = (FA)\frac{Sp}{1-Sp} \tag{A.11}$$

$$FA = (CR)\frac{1-Sp}{Sp} \tag{A.12}$$

Finally, starting with equation (1), multiplying through by $(H+M)$ yields:

$$Sn\,(H+M) \;=\; H \tag{A.13}$$

or $\qquad H\,(1-Sn) = Sn\,(M) \tag{A.14}$

which gives rise to the following relationships:

$$H = (M)\frac{Sn}{1-Sn} \tag{A.15}$$

$$M = (H)\frac{1-Sn}{Sn} \tag{A.16}$$

To now develop the dependency between *PPP*, *NPP*, *Sn* and *Sp*, set CR=CR using equations (A.3) and (A.11) yielding:

$$M\left(\frac{NPP}{1-NPP}\right) = FA\left(\frac{Sp}{1-Sp}\right) \tag{A.17}$$

Now set M=M in equations (A.4) and (A.16) which results in:

$$CR\left(\frac{1-NPP}{NPP}\right) = H\left(\frac{1-Sn}{Sn}\right) \tag{A.18}$$

Following the same procedure by setting H=H in equations (A.7) and (A.15) which yields:

$$FA\left(\frac{PPP}{1-PPP}\right) = M\left(\frac{Sn}{1-Sn}\right) \tag{A.19}$$

Finally setting FA=FA in equations (A.8) and (A.12) results in the following relationship:

$$H\left(\frac{1-PPP}{PPP}\right) = CR\left(\frac{1-Sp}{Sp}\right) \tag{A.20}$$

The relationships (A.17)-(A.20) can be further reduced. Computing the relationship (CR)/H from both equations (A.18) and (A.20) yields:

$$\frac{CR}{H} = \frac{\left(\dfrac{1-Sn}{Sn}\right)}{\left(\dfrac{1-NPP}{NPP}\right)} = \frac{\left(\dfrac{1-PPP}{PPP}\right)}{\left(\dfrac{1-Sp}{Sp}\right)} \tag{A.21}$$

Cross multiplying implies the following classical dependence between *PPP*, *NPP*, *Sn* and *Sp:*

$$\left(\frac{1-Sn}{Sn}\right)\left(\frac{1-Sp}{Sp}\right) = \left(\frac{1-PPP}{PPP}\right)\left(\frac{1-NPP}{NPP}\right) \tag{A.22}$$

To check why this is true, similar calculations can be made using both equations (A.17) and (A.19) for (FA)/M:

$$\frac{FA}{M} = \frac{\left(\dfrac{Sn}{1-Sn}\right)}{\left(\dfrac{PPP}{1-PPP}\right)} = \frac{\left(\dfrac{NPP}{1-NPP}\right)}{\left(\dfrac{Sp}{1-Sp}\right)} \tag{A.23}$$

Cross multiplying implies the reciprocal of equation (A.22) is true which further validates this approach because if things are equal, their reciprocals should also equate:

$$\left(\frac{Sn}{1-Sn}\right)\left(\frac{Sp}{1-Sp}\right) = \left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right) \tag{A.24}$$

Appendix B – The Wald Algorithm for *PPP* and *NPP*

Starting with the Wald test of equation (11), where α and β are known or measured (possibly time varying).

$$\log\frac{\beta}{1-\alpha} \;\le\; \gamma(t) \;\le\; \log\frac{1-\beta}{\alpha} \tag{B.1}$$

where, from equations (6) and (8)

$$\alpha = 1 - Sp \tag{B.2}$$

and $\qquad \beta = 1 - Sn \tag{B.3}$

Thus the Wald optimality criterion (B.1) can be written:

$$\log\frac{1-Sn}{Sp} \;\le\; \gamma(t) \;\le\; \log\frac{Sn}{1-Sp} \tag{B.4}$$

But it is known that

$$\left(\frac{Sn}{1-Sn}\right)\left(\frac{Sp}{1-Sp}\right) = \left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right) \tag{B.5}$$

Thus $\quad \left(\dfrac{1-Sn}{Sp}\right) = \left(\dfrac{Sn}{1-Sp}\right)\left(\dfrac{1-PPP}{PPP}\right)\left(\dfrac{1-NPP}{NPP}\right) \tag{B.6}$

Hence $\left(\dfrac{1-Sn}{Sp}\right) = \left(\dfrac{1-\beta}{\alpha}\right)\left(\dfrac{1-PPP}{PPP}\right)\left(\dfrac{1-NPP}{NPP}\right) \tag{B.7}$

Also starting with equation (B.5) yields:

$$\left(\frac{Sn}{1-Sp}\right) = \left(\frac{1-Sn}{Sp}\right)\left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right) \tag{B.8}$$

Which can be written as:

$$\left(\frac{Sn}{1-Sp}\right) = \left(\frac{\beta}{1-\alpha}\right)\left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right) \tag{B.9}$$

Hence the optimal Wald sequential algorithm can be modified to include the *PPP* and *NPP* terms as follows: For the test statistic $\gamma(t)$ assuming constant α and β values with

5

NPP and *PPP* possibly changing with time, the new regions for testing now become:

$$\log\left[\left(\frac{1-\beta}{\alpha}\right)\left(\frac{1-PPP}{PPP}\right)\left(\frac{1-NPP}{NPP}\right)\right] \leq \gamma(t)$$

$$\leq \log\left[\left(\frac{\beta}{1-\alpha}\right)\left(\frac{PPP}{1-PPP}\right)\left(\frac{NPP}{1-NPP}\right)\right] \qquad \text{(B.10)}$$

Updated Studies in Approximate Entropy involving Fractional Noise and Fatigue Data

D. W. Repperger[1], R. A. McKinley[1], K. A. Farris[1], J. S. Warm[1], C. M. Walters[2], and B. A. Schmitt[2]

[1-] 711 HPW AFRL/RH, WPAFB, Ohio 45433
[2-] Wright State University, Dayton, Ohio 45435

Abstract

New applications of approximate entropy are currently ongoing and being analyzed with innovative and novel data being generated within AFRL. This paper reviews some extant studies in the area of Human Machine studies and describes the basics of this powerful use of an information theoretic measure. The goal is to detect if the state of a human may have been compromised by fatigue, or other stressors, utilizing noninvasive sensory measurements.

I. Introduction

New investigations at the 711 HPW WPAFB have been focused on the use of a measure of uncertainty in data termed "approximate entropy." This new interest in research in the USAF involves studies of humans under fatigue induced by loss of sleep [1] or other stressors including emotional stress [2]. Herein will be presented some of the basics of the concepts underlying the calculation of approximate entropy. Prior work in this area [3] has considered determining when loss of consciousness would occur during high acceleration stress. The study in [3] was based on an exogenous physical stress (G or acceleration forces) on a human. In [4], however, a mental trauma involving cognitive stress (a high workload task) was also investigated using the approximate entropy measure. Periods of high and low task difficulty could be gleaned from the performance of the subjects and the approximate entropy measure.

II. Principles of Approximate Entropy

The concepts involving approximate entropy are simple. In Figure (1) a signal $s(t) = s_1(t)$ is compared to itself displaced one sample $s(t+\Delta t) = s_2(t)$. The entropy is the normalized amount of disorder between the two adjacent signals $s_1(t)$ versus $s_2(t)$ or $s(t)$ versus $s(t+\Delta t)$. If the approximate entropy is low, this disorder is minimal. If this disorder or variation is high, then the normalized entropy may reach values of 1.0 or larger. In references [3] and [4] more technical details are included. Appendix A includes

7

reference [3] involving physical stress on humans for informative purposes. Appendix B includes reference [4] with a cognitive stress investigation.
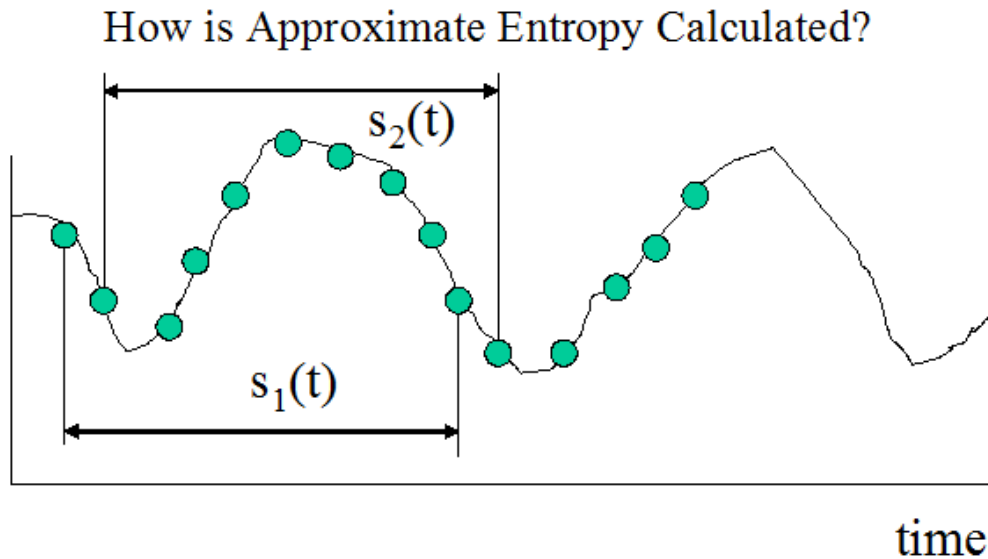
Figure (1) – Determination of the Approximate Entropy Measure

References

[1] R. A. McKinley, R. Schmidt, D. W. Repperger, A. Pinchak, B. Yu, L. McIntire, and M. Kane, "Evaluation of the Eye-Com Alertness Monitoring Device for Military Aviation Applications," AFRL-HE-WP-TR-2008-XXXX.

[2] R. M. Schmidt, "Facial Expressions of Emotion and the Assessment of Performance," Protocol F-WR-2009-0008-H, 2009.

[3] D. W. Repperger, W. B. Albery, and L. D. Tripp, "Approximate Entropy as an Assessment Tool for System Complexity and Performance Valuation in Human-Machine Systems," *Proceedings of the 9th IFAC Symposium on Human-Machine Systems*, Sept 7-9, 2004, Georgia Tech., Atlanta, Georgia.

[4] D. W. Repperger, and J. J. Skelly, "A Real-Time Measure to Study Dynamic Interactions with A Visual Display," *Proceedings of the 9th IFAC Symposium on Human-Machine Systems*, Sept 7-9, 2004, Georgia Tech., Atlanta, Georgia.

Appendix A – Reference [3] included for information purposes

APPROXIMATE ENTROPY-AN ASSESSMENT TOOL FOR SYSTEM COMPLEXITY AND UNCERTAINTY

D. W. Repperger[1], W. B. Albery[1], L. D. Tripp[2]

[1] Air Force Research Laboratory, AFRL, WPAFB, Ohio 45433, USA
[2] General Dynamics, Dayton, Ohio, 45433, USA

Abstract: For complex systems and interactions, a means of determining the quality of information in a time series involving data generated from a human-machine system is investigated. This recent numerical analysis method, widely accepted in the medical field, termed "Approximate Entropy" can provide a compelling means of uncovering irregularity in data of all types. Changes in the state of a human-machine system can be quickly gleaned in real time, on line. Application of this measure is investigated on tracking performance data when the human-machine system is known to be compromised in a performance sense.

Keywords: Computational methods, classifiers, entropy, discriminations, dynamic behavior.

1. INTRODUCTION

Many complex Human-Machine systems may have a sudden change in state and it is desired to detect, in real time, whether a measured or derived signal may now differ from its prior situation. Traditional statistical measures (mean, standard deviation or other moments) cannot always capture the change in some signals because of the complexity of the underlying system dynamics. Pioneered by Pincus (1991-2), discussed in Pincus and Kalman (1997) and in numerous publications, an interesting new measure has been developed to help detect changes in the "regularity" of a time series termed "Approximate Entropy" (ApEn). By comparing a time series with itself, over time, this measure of system irregularity has found applicability in the medical field for discerning differences in heart beat either due to sleep cycle or disease, Yeragani, et al. (1998) and Pincus (1992). This procedure has applicability in distinguishing hormonal changes, Gevers, et al. (1998), mood swings, Pincus (2003), and in other complex applications involving very intricate signals. Other applications include EMG and tremor distinction (Morrison and Newell, 2000), EEGs (Bruhn et al., 2000), recognizing epileptic activity (Diambra, et al. 1999) and for discerning cocaine addiction (Newlin et al., 2000). Generalizing this concept even further, Sugihar (1990), it is now possible to determine if the underlying complexity of a system state may have changed when viewed within the context of a control system by observing some key output data. This methodology has particular interest in identifying if the underlying dynamics of a process may or may not exhibit chaotic behaviour and to predict this possibility.

Approximate entropy is based on the simple principle that if a time series signal can be compared to itself (heart beat data, for example) and the amount of the disorder in a comparison or change

9

(entropy) between relative time shifts of these data is increasing, this is probably an indication of some change in state. This differs from traditional correlation measures, since they are not based on information theory concepts and may require fixed implicit models. The use of ApEn is model independent and only depends on the real time data series. There are at least four reasons why approximate entropy provides new information about system complexity not normally derived from typical statistical measures (first and second order moments):

(1) If data are noisy, the approximate entropy measure can be compared to the noise level in the data to determine what quality of true information may be present in the data.

(2) If the data have an artifact, this does not impact the approximate entropy measure as much as it would affect typical first and second order statistical moments from the data.

(3) Approximate entropy can be designed to work for small data samples (n < 50 points) and can be applied in real time, on line. Thus changes in the state of a physical process may be quickly determined.

(4) For pure stochastic processes, approximate entropy will become practically infinite.

Thus, the quality of the information in a signal can then be quantitatively evaluated by comparing the entropy level of the measured signal with its underlying (non random) signal component.

To test the concept of approximate entropy on performance data, an extensive database exists involving pilots performing cognitive tasks while simultaneously being subjected to high acceleration stress, Tripp (2001). The volunteer subjects were required to perform both a math computation task as well as manual tracking in a position-control mode. Data were recorded prior to the event of loss of consciousness, during the period the subjects went unconscious, and afterwards in the post recovery period. The approximate entropy measure was calculated on the tracking data to quantify, objectively, the amount of disorder in the subject's response during different time epochs of the experiment.

2. EXPERIMENTAL SCENARIO

At the Air Force Research Laboratory, WPAFB, Ohio, the data presented here partially involved a three axis motion simulator that is used to determine a pilot's response to acceleration stress. The system shown in Figure 1 has a 5.8 meter radius with a large spherical cab and can create a force of 20 G at a rotational velocity of 56 RPM. Such a system weights 163,000 kilograms.
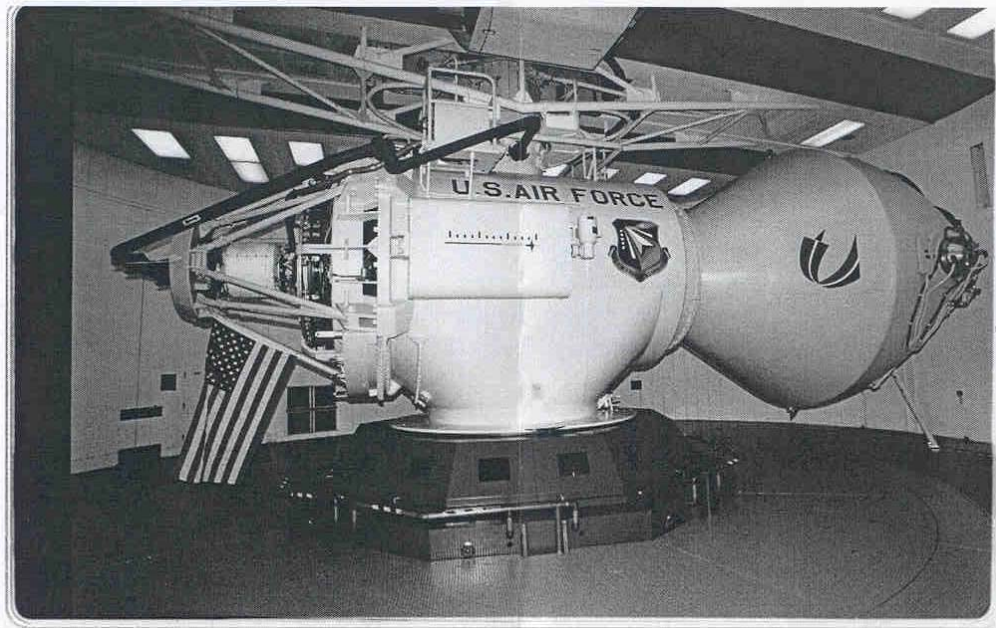
10

Figure 1 – The Dynamic Environment Simulator to Produce Acceleration Stress

3. OBJECTIVE

The objective is to test if the measure of uncertainty or irregularity known as approximate entropy can be a valuable tool in assessing whether a pilot is compromised in a cognitive sense. The underlying assumption is that under high acceleration stress conditions, human data are extremely more variable and thus should show higher entropy. In addition, if the time rate of increase of entropy is positive going into the loss of consciousness event and negative coming out of the loss of consciousness event, then the rate of change of approximate entropy may be a valuable prediction tool to ascertain shifting cognitive state. Also, if the rate of change of approximate entropy is decreasing, this may be a valuable indicator that the mission capability of the pilot has returned to a normal level and the effect of the untoward event (unconsciousness) may no longer impact the mission effectiveness of the pilot.

4. HYPOTHESIS

The null hypothesis that we wish to reject is that the approximate entropy metrics (magnitude and time rate of change of ApEn) will not significantly vary during known changes of the cognitive state of the pilot being stressed. Only the performance tracking data will be used to ascertain the cognitive state of the pilot.

5. METHODS

Data from 16 USAF pilots/subjects were collected as they performed a compensatory tracking task and carried out a math computation exercise in a large centrifuge motion simulator. These subjects performed these tasks under high acceleration stress. The acceleration stress increased until the

11

subjects became unconscious (GLOC or G loss of consciousness). There were 3 or 4 data days for each subject.

6. APPARATUS

Two centrifuges were used in this study. The DES centrifuge at WPAFB was described in section 2 of this paper. A second centrifuge at Brooks Air Force Base, Texas was employed to generate similar acceleration profiles. The same performance tasks (math computation and manual tracking) were used with both motion simulators on qualified human subjects.

7. EXPERIMENTAL DESIGN

Both genders of healthy USAF pilots/subjects participated in this experiment. The end point condition was the loss of consciousness. A recovery period followed the GLOC event.

8. RESULTS

Figure 2 shows the stressor (lower plot - G acceleration level versus time) and the root mean square tracking error (upper plot in Figure 2) for one subject on the first of his four days when data were collected. The events of pre-GLOC (prior to G loss of consciousness), the period of complete incapacitation, and the post recovery period are indicated. Figures 3a-b is a similar plot of the approximate entropy function versus time with some of these same event markers noted based on 10 second intervals of the data. One observes in Figure 2, from the performance data (top plot of Figure 2), that prior to the GLOC event, a pilot's behaviour was manifested by a high variation in the tracking error signal. Coming out of the GLOC event, in the same diagram, high variability in the tracking error also exists for 40 or more additional seconds. Eventually the data show a reduction in the variability (t > 100 seconds) as the pilot returns to a cognitive state, typical of normal tracking. Hence, disorder (high levels of approximate entropy) in the tracking error signal seems related to a compromised cognitive state of the pilot.
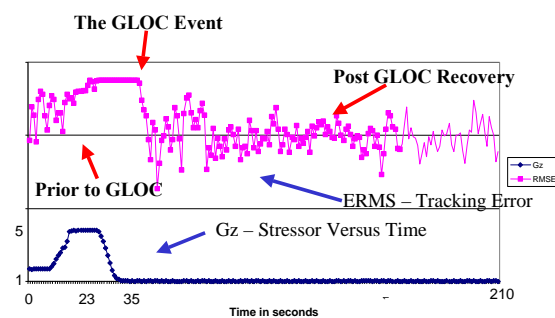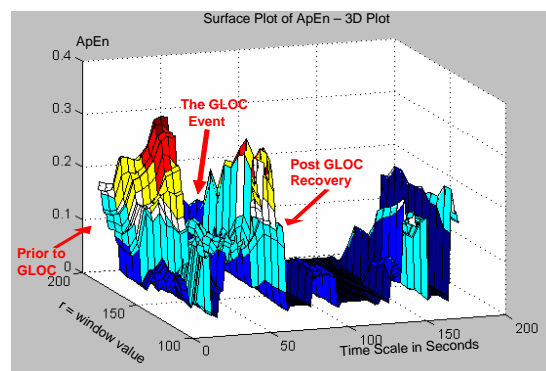

Figure 2 – ERMS and G-Stress Versus Time

Figure 3a – Subject 1 – Day 1 – 3D Surface Plot of Approximate Entropy
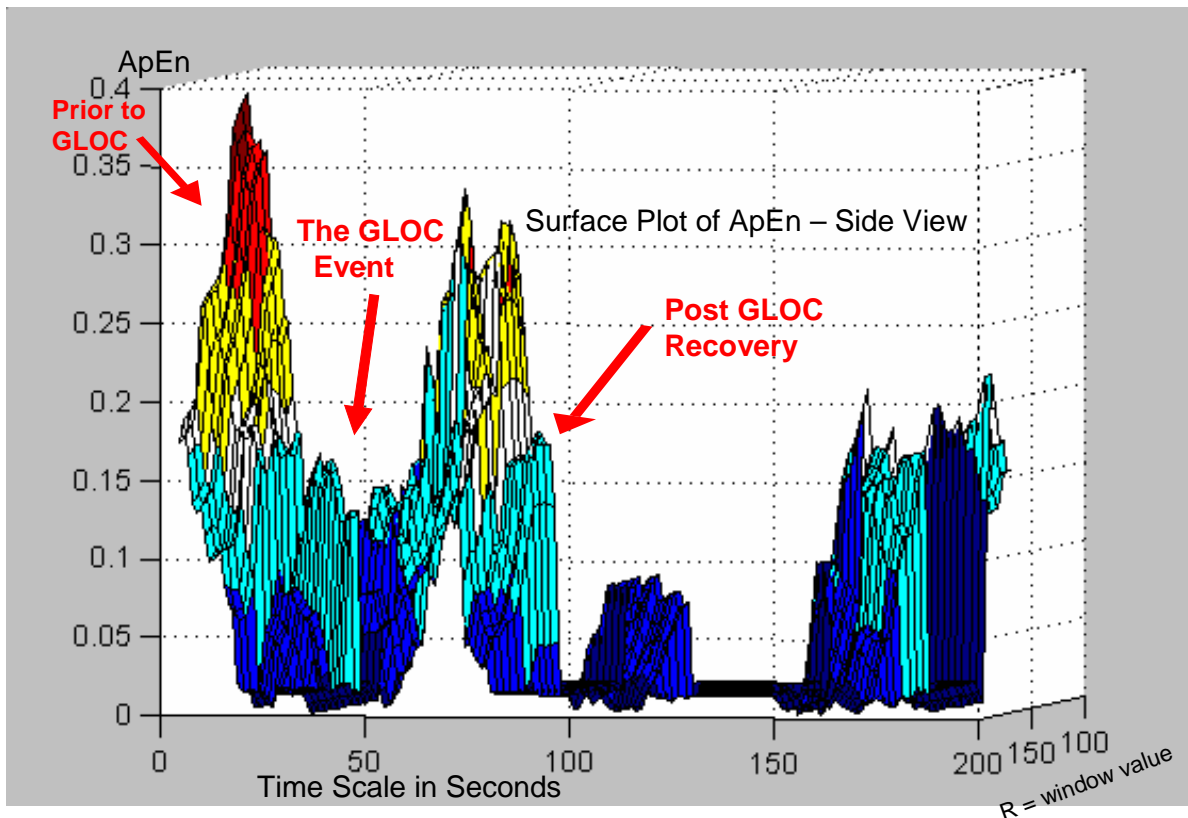
12

Figure 3b – Subject 1 – Day 1 – Side View of Approximate Entropy

The appendix describes, in a succinct manner, important technical aspects of this measure, approximate entropy, as compared to alternative metrics involving time series data. With reference to Figures 3a-b, the approximate entropy ApEn is plotted versus time on the x axis. One thing is clear, i.e. prior to the GLOC event and during the post GLOC event, that the absolute value of ApEn and the magnitude of the rate of change of ApEn with respect to time are extremely high. In Figures 3a-b, the almost zero value of ApEn near the fifth time epoch (t = 50 seconds) needs to be explained. Referring to Figure 2, during the time period (t ϵ [25,35]) seconds, the RMS tracking error is constant (worst case) because the subject is unconscious. There is little variation in his relative behavior (as measured by tracking performance, which is not changing), hence ApEn $\rightarrow$ 0 in this interval. Therefore to use a measure such as ApEn to predict or indicate the likelihood of the pilot to become unconscious, one must take into account the magnitude of ApEn as well as it's time rate of change. Figure 4 depicts a majority voting classifier system that could use this information to predict the cognitive state of the pilot using as inputs ApEn measures and tracking performance. Since brevity must be the style in this paper, we only report on the data analysis measure of ApEn with respect to the known cognitive state of the pilot.

13

RMS
(Root
Mean
Square)
Tracking
error (t)

ApEn(t)

(d/dt)ApEn(t)

+

+

+

Majority

Voting

Classifier

Pilot is conscious

Pilot is unconscious
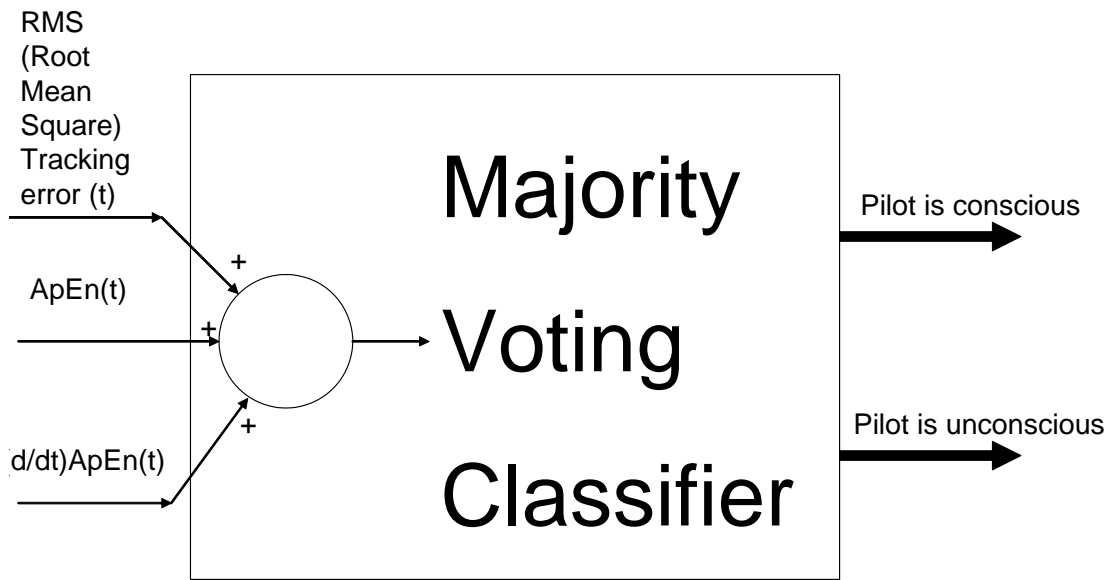
Figure 4 – A Majority Voting Scheme to Predict the Pilot's Cognitive State

Table 1 illustrates the mean (μ) and standard deviation (SD) values of ApEn and its time derivative fifteen seconds prior to the GLOC event. Also in Table 1 are these respective values 100 seconds post GLOC for comparison purposes. The data are averaged across 16 subjects with 3-4 data days for each subject. Figure 5 illustrates these mean and SD values averaged over the 16 subjects and for those data days with pre GLOC data at 15 or more seconds prior to the incapacitation event.

Table 1 –ApEn and its time derivative 15 seconds Pre GLOC and 100 seconds Post GLOC

Pre GLOC				Post GLOC			
ApEn		(d/dt) ApEn		ApEn		(d/dt) ApEn	
μ	SD	μ	SD	μ	SD	μ	SD
.43	.07	.013	.006	.32	.13	.01	.006

Table 2, displays the results of a one-way ANOVA, one factor for Magnitude of ApEn and Table 3 displays these values for the magnitude of the time rate of change of ApEn. Please note the time samples are one second apart for the data displayed in Figure 2. Figure 6 shows the results of the JMP (developed by the SAS Institute) analysis for comparison of means (Tukey-Kramer test).

14
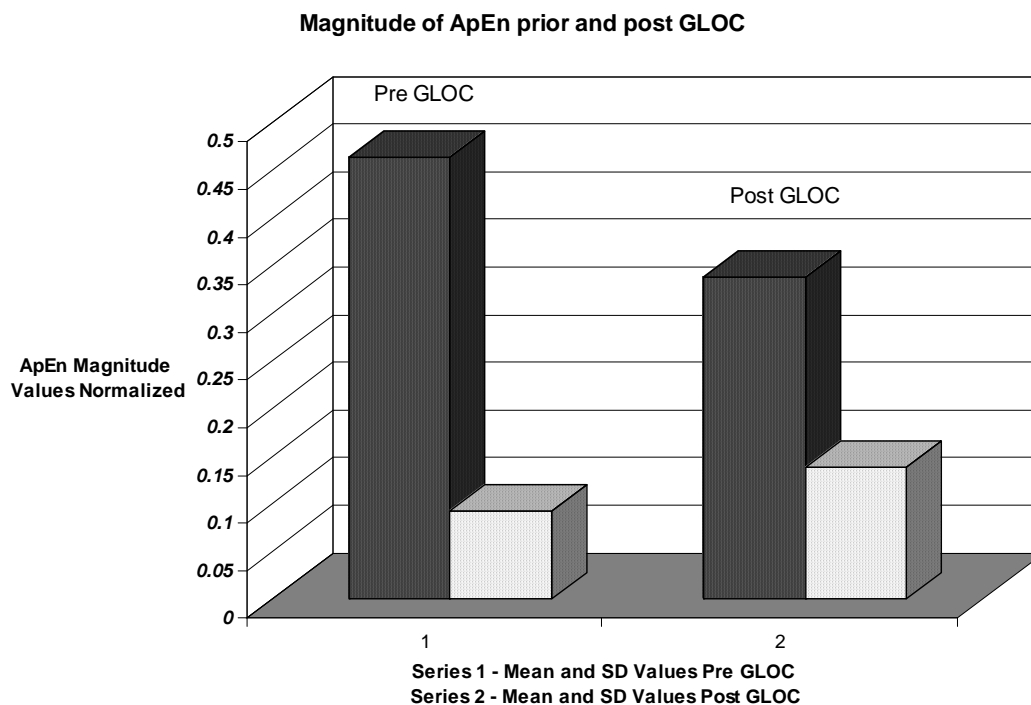
Magnitude of ApEn prior and post GLOC

Figure 5a – Magnitude of ApEn for Pre versus Post GLOC
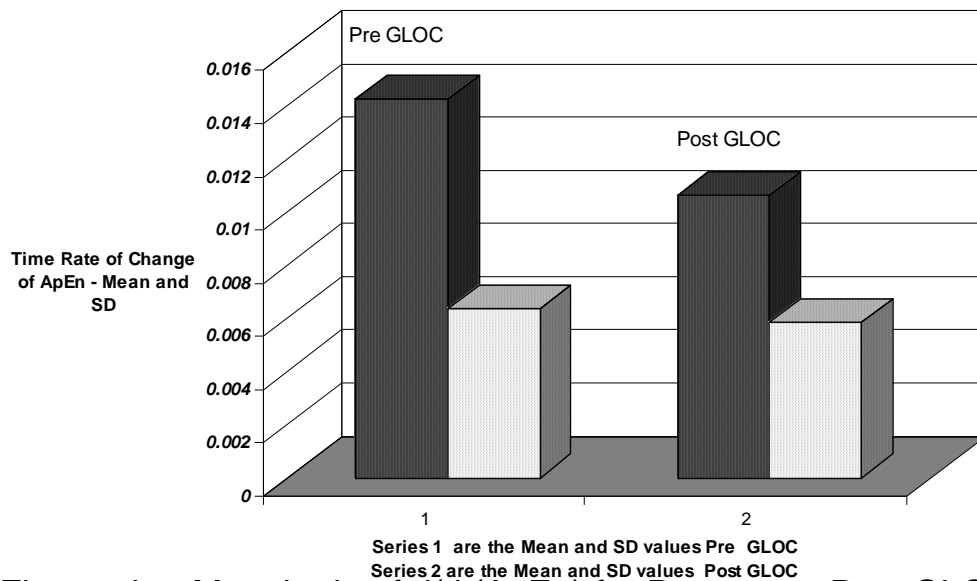
Figure 5b – Magnitude of d/dt(ApEn) for Pre versus Post GLOC

Table 2 – One Way ANOVA for Mag. of ApEm

Source	DoF	Σ Squares	Mean Squares	F Ratio	Prob>F
GLOC	1	0.2887	0.2887	20.7408	<.0001
Error	70	0.9742	0.0139		
C. Total	71	1.2629			

15

Table 3 – One Way ANOVA for Mag of d/dt (ApEn)					
Source	DoF	Σ Squares	Mean Squares	F Ratio	Prob>F
GLOC	1	0.00024	0.00024	6.3978	.0137
Error	70	0.0026	0.000037		
C. Total	71	0.00284			
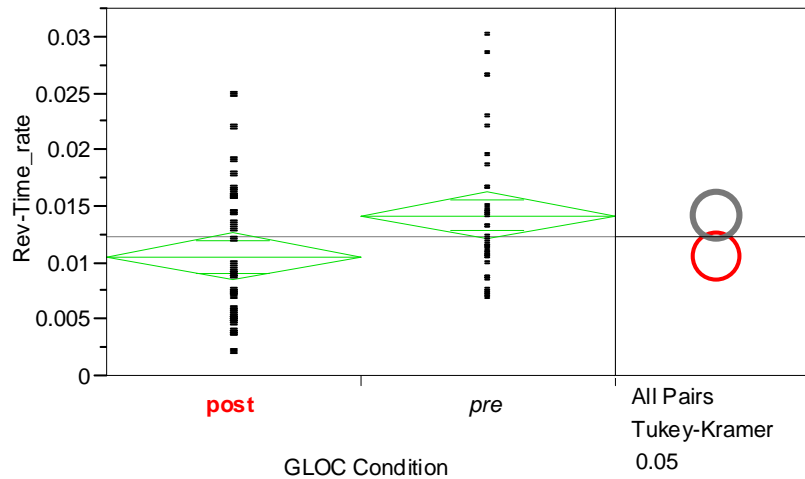
Figure 6 – JMP Analysis for d/dt (ApEn) Data – Pre and Post GLOC.

9. DISCUSSION

Preliminary results indicate that prior to GLOC the approximate entropy function is both high and shows increasing positive rates with respect to time. In the recovery period, the approximate entropy function seems to show decreasing rates as the pilot's variability in his performance seems to be more reduced. The overall results of this study across all 16 subjects with three to four days of testing for each subject are reported.

10.   CONCLUSION

A study of the efficacy of using approximate entropy to predict the cognitive state of the pilot is conducted on human subjects known to be compromised in a performance sense due to loss of consciousness. Studies are ongoing to predict the loss of consciousness prior to the event by examining the rate of change of approximate entropy in real time, on line, using other signals such as physiologically based data including oxygen perfusion to the subject's brain, etc.

11.   REFERENCES to [3]

Bruhn, et al. (2000). "Electroencephalogram Approximate Entropy Correctly Classifies the Occurrence of Burst   Suppression Pattern as Increasing Aesthetic Drug Effect," *Anesthesiology*, **93**, 4, Oct, pp. 981- 985.

Diambra, L., Bastos de Figueiredo, J. C., and Malta, C. P.(1999). Epileptic activity recognition in EEG recording, *Physica A,* **273**, pp. 495-505.

Gevers, E., Pincus, S. M., Iaian, C. A., Robinson, A. F., and Veldhuis, J. D. (1998). "Differential Orderliness of the GH release Process in Castrate Male and Female Rats," *The American Journal of Physiology*, pp. R437- R444.

Hilborn, R. C. (1994). *Chaos and Nonlinear Dynamics*, Oxford University Press.

Kolmogorov, A. N.(1958). A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk SSSR* **119**: 861-864.

Morrison, S., and Newell, K. M. (2000). "Postural and resting tremor in the upper limb, *Clinical Neurophysiology*, **111,** pp. 651-663.

Newlin, D.B., et al. (2000). "Intravenous Cocaine Decreases Cardiac Vagal Tone, Vagal Index (Derived in Lorenz Space), and Heart Period Complexity (Approximate Entropy) in Cocaine Abusers, N*europsychopharmacology*, **23**,5, pp. 560- 568.

Pincus, S. M. (1991). "Approximate Entropy as a measure of system complexity," *Proc. Natl. Acad. Sci,* USA, **88**, pp. 2297-2301.

Pincus, S. M., and Viscarello, R.R. (1992). "Approximate Entropy: A Regularity Measure for Fetal Heart rate Analysis," *Obstetrics and Gynecology*, **79**, (2), February, pp. 249- 255.

Pincus, S. M. and Kalman, R. E. (1997). "Not all (possibly) "random" sequences are created equal," *Proc. Natl. Acad. Sci,* USA, **94**, pp. 3513-3518.

Pincus, S. M. (2003). "Quantitative Assessment Strategies and Issues for Mood and Other Psychiatric Serial Study Data," *Bipolar Disorders*, **5**, pp. 287-295.

Sugihar, G. and R. M. May (1990). "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, **344**, April 19, pp. 734-741.

Tripp, L. D. (2001).   "Enhanced Recovery of Aircrew from Acceleration Induced Loss of Consciousness (GLOC)," Veridian Engineering, Inc., October.

Williams, G. P. (1997). *Chaos Theory Tamed*, Joseph Henry Press).

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). "Determining Lyapunov Exponents from a Time Series," *Physica 16D*, pp. 285-317.

Yeragani, V. K., et al. (1998). Fractal dimension and approximate entropy of heart period and heart rate: awake versus sleep differences and methodological issues," *Clinical Science*, **95**, pp. 295-301.

APPENDIX from [3] – TECHICAL DETAILS REGARDING APPROXIMATE ENTROPY

Approximate entropy is sometimes termed a "regularity measure" which quantifies the unpredictability of fluctuations in a time series, e.g. in instantaneous heart rate signal. High levels of ApEn (highly irregular) reflect the likelihood that "similar" patterns of observations will NOT be followed by additional "similar" observations. Thus a more complex process has high levels of ApEn and small values of ApEn imply predictable (repetitive) patterns are inherent in the data. Uncertainty (or high system complexity) is related to high levels of ApEn.

Conversely, low values of ApEn indicate predictability of a time series. Given a series of s(t) measurements (for this case n < 50) s(1), s(2), … s(n), equally spaced in time, the ApEn of this data series depends on two key parameters: m and r.   m is an integer that represents the length of compared runs (a window or how many data samples the two series differ) and r, effectively, represents a filter. Typically m = 1 or 2 which distinguishes the two series and r is a tolerance measure (criterion of similarity). The next steps provide the procedure to compute the ApEn versus time:

<u>Step 1:</u> Form a sequence of vectors x(1), x(2), .., x(n), where each x(i)=[s(i), s(i+1), …,s(i+m-1)].

<u>Step 2:</u> Use the sequence x(1), x(2), …, x(n) to construct, for each i, 1<i<n-m+1, C_i^m(r) = (the number of x(j) such that d[x(i), x(j)] < r) / (n-m-1). The distance metric d satisfies:

$$d[x(i),x(j)] = max \ |s(i+k-1) - s(j+k-1)| \qquad\qquad (A.1)$$
$$for \ k=1,2,…,m$$

Hence d represents the distance between the vectors x(i) and x(j), given by the maximum of their respective scalar components. The next step defines the logarimithic entropy measure:

$$\underline{Step\ 3:} \qquad\qquad\qquad n-m+1$$
$$\phi^m(r) = (n-m+1)^{-1} \ \Sigma \ ln \ C_i^m(r) \qquad\qquad (A.2)$$
$$i=1$$
$$resulting\ in$$
$$ApEn = \phi^m(r) - \phi^{m+1}(r) \qquad\qquad (A.3)$$

We now show a relationship of ApEn to some of the objective means of quantifying chaos in nonlinear dynamical systems.   Three popular means of quantifying chaos are briefly discussed here which include (1) Lyapunov Exponents, (2) Embedding Dimensions, and (3) K-S Entropy. Large values of ApEn can be related, e.g., to a positive Lyapunov exponent.

(1) <u>Lyapunov Exponents:</u>
In a chaotic system, the Lyapunov Exponent represents the exponential rate of separation of adjacent trajectories.   Consider a nonlinear system near a fixed point (x_0) where:

$$d/dt \ x(t) = f(x(t)) \qquad\qquad (A.4)$$

Taking the Taylor series expansion near x_0 results in:

$$d/dt \ x(t)= 0 = f(x) = f(x_0) + (x-x_0) \ df/dx +... \qquad\qquad (A.5)$$

Define a new variable $z=x-x_0$. Then proximal to the fixed point: d/dt z(t) = z [df/dx] |x_0 which has solution: z(t) = z(0) e $^{\lambda t}$, where the Lyapunov exponent λ = [df/dx]|x_0 is a characteristic value of the fixed point. Thus two adjacent trajectories are attracted to each other if they approach the fixed point and λ <0. However, if two adjacent trajectories repel each other, then λ > 0 and this is typical of chaotic behaviour. Thus if the average Lyapunov exponent is positive near a fixed point, this is a viable definition of chaotic behavior since nearby trajectories have exponential divergence in phase space (Hilborn, 1994).

(2)<u>Embedding Dimensions:</u> The embedding dimension is the number of points necessary to predict the next point in the time series and is an objective measure of system complexity. Thus highly complex systems would have a high measure of embedding dimension and this may indicate the potential for chaos (Williams, 1997).

<u>For K-S Entropy:</u> The Kolomogorov-Sinai entropy (Kolmogorov, 1958) deals with the probability that a given trajectory point falls within some particular region of state space. Thus a chaotic system

18

would have a high entropy measure since it may have a nonzero high level of probability of accessing different areas of the state space. Pincus (1991) has shown the distinction between ApEn and K-S entropy by noting that the K-S entropy measure (a theoretical construct) can be defined via:

For K-S Entropy:

$$\lim_{r \to 0} \quad \lim_{m \to \infty} \quad \lim_{n \to \infty} [\phi^m(r) - \phi^{m+1}(r)] = \Delta t\, h(\rho) \tag{A.6}$$

where $h(\rho)$ is the K-S entropy. Choosing $\Delta t = 1$ shows the relationship and distinction between ApEn and the K-S entropy measure. In other words, the K-S entropy is a theoretical construct which requires r to be near zero, and m and n to be large in value. ApEn, on the other hand, may have r nonzero (but small) and finite values of m and n. Thus ApEn is more applicable to real world situations and data that are collected in an experimental scenario. Other methods also exist for quantifying chaos include fractal dimension and correlation dimension (Hilborn, 1994).

Finally, we describe, and show by example, how the calculation of ApEn can be conducted online in real time for data consisting of finite samples. To physically understand the key parameters r and m, Figures 7a-b displays a time series, where s(t) is the signal of interest. The signal s(t) has two constituent signals $s_1(t)$ and $s_2(t)$ derived by shifting the data one data sample (m=1) over a 12 sample time epoch (n=12). $s_2(t)$ is just one data sample (m=1) shifted to the right from $s_1(t)$. Hence, the tolerance variable r would be gleaned from the difference between $s_1(t)$ and $s_2(t)$ as indicated in Figure 7b (plotted on different and exaggerated vertical scales for illustrative purposes). Thus the approximate entropy value depends heavily on the n, m, and r values as well as the regularity of the signal s(t) over the interval in which it is being examined.

In conclusion, a relationship between the Lyapunov exponent and the prior procedure using ApEn shows a further hidden connection of the proposed method to nonlinear dynamics. One can derive the Lyapunov exponent from real data (Wolf, et al., 1985) and it is desired to distinguish the situation of chaos from possible measurement error in a time series (Sugihar and May, 1990) using real time measurements. To show how to employ the Lyapunov exponent method with real data, the assumption is first made that the data are sampled equally in time. The data samples $s(t_0)$, $s(t_1),s(t_2),\dots,s(t_n)$ are labelled s_0, s_1, s_2, $\dots$, s_n. If τ is the constant time interval between samples, then for some integer n, the following relationships exists:

$$t_n - t_0 = n\ \tau \tag{A.7}$$

Now a system will behave chaotically if the divergence of nearby trajectories are showing exponential changes in their differences, i.e.

$$d_0 = |\, x_j - x_i| \tag{A.8}$$

$$d_1 = |\, x_{j+1} - x_{i+1}| \tag{A.9}$$

$$d_2 = |\, x_{j+2} - x_{i+2}| \tag{A.10}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

19

$$d_n = | x_{j+n} - x_{i+n}| \qquad\qquad (A.11)$$
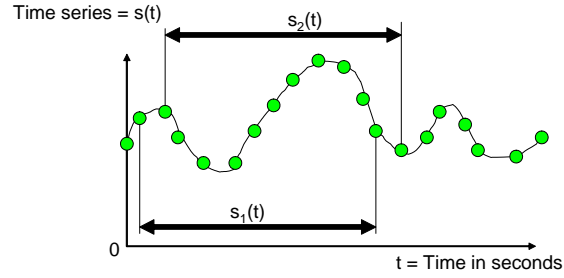
Figure 7a – Original Signal s(t) to be analyzed versus time.
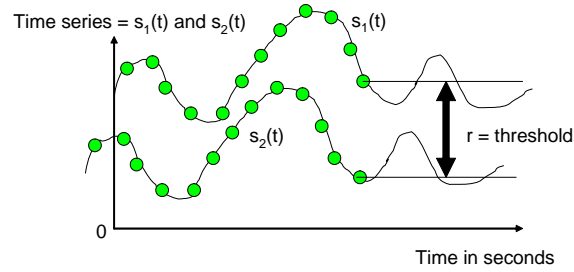
Figure 7b – Constituent signals $s_1(t)$ and $s_2(t)$ derived from s(t).

is assumed to exponentially increase (on the average) as n gets larger. The assumption is really that:

$$d_n = d_0\ e^{\lambda n} \qquad\qquad (A.12)$$

and the quantity λ can be determined via

$$\lambda =\ (1/n)\ \ln\ (d_n/d_0) \qquad\qquad (A.13)$$

and if λ is positive, then the behavior would be chaotic. A graphical means to determine this effect is to make a log plot of the respective difference ratios (d_n/d_0) and determine if the line that best fits the data has a nonzero slope.

If the slope of the line is nonzero (and statistically different from zero) then $\lambda \neq 0$ and the data either represent a chaotic system ($\lambda>0$) or an attractor ($\lambda<0$), the latter of which would indicate convergence to a fixed point. Hence λ is similar to the Lyapunov exponent.

20

- -

A Real-Time Measure to Study Dynamic Interactions with a Visual Display

D. W. Repperger and J. J. Skelly

Abstract: As operators interrelate with complex systems, such as those that occur in flying or operating aircraft, the true performance and interaction is of a dynamic nature. However, most analyses of human interface systems employ the assumption that the display is static and does not change with time. This paper will consider a real time measure (approximate entropy) to study and evaluate human-machine interaction as events in the environment change in a dynamic sense. The experimental paradigm to evaluate the efficacy of the proposed real time measure will be borrowed from studies investigating the effects of spatio-temporal structure on cognition.

Keywords: Dynamic behavior, entropy, man-machine interfaces, nonstationary.

1. INTRODUCTION

In many physical systems, the presumption that a static environment captures the essence of the performance of the human-machine interaction is extremely limiting. This is equivalent to the assumption of linearity. In linear systems, the transient response has commonalty with the steady state response through key parameters, such as eigenvalues. However, for nonlinear systems, this is not the case and most human interactions with complex systems are more typically nonlinear. In a companion paper, Repperger, et al. (2004), a real-time measure of entropy or disorder has been employed to analyze data from an experiment when it was known that pilots became unconscious and their behaviour changed dramatically. The real time measure of approximate entropy (ApEn - Pincus and Viscarello (1992), Gevers et al. (1998), and Pincus (2003)) has been embraced by the medical and psychological communities to show that certain changes in real time data can indicate an alteration in the medical or psychological state of a patient. This is not unlike humans interacting with dynamical systems, when they become overloaded and task performance changes, accordingly. This study will consider the real time measurement of irregularity in data (approximate entropy) as applied to key performance parameters from a spatial-temporal investigation on human perception. It is known that, in scenarios of the data to be presented, human performance can be compromised by the procedure at which information is displayed to the operator. This is true for two, almost identical performance tasks to be offered, where the complexity of the task (bits) and rate of presentation (bits/second) are identical for both tasks, but their dynamic attending characteristics differ. From a Fitts' Law perspective (Fitts, 1954) the two tasks provide identical difficulty but it will be shown in the sequel that the performance of the subjects is markedly dissimilar. The only distinction between the tasks is that they require different dynamic attention resources from the human subjects. This experimental platform is ideal for evaluation of the ApEn metric which may be able to discern differences in operator responses, when Fitts' law would not be able to make such a fine distinction.

A powerful paradigm to investigate the effects of dynamic information structure on human performance can be seen in the works of Skelly and colleagues (Skelly (2003), Skelly et al. (2000), and Jones and Skelly (1993)). To briefly summarize the results of some of their studies relevant to this investigation, the environment of an operator is made to change both in a temporal and/or spatial sense with respect to a visual display. The timing and spatial display of this information to the operator may affect a viewer's ability to allocate attention among items within a single stream of information events or among multiple information streams. Hence these spatio-temporal structural properties may serve to facilitate or interfere with attending to the specific arrival time or location associated with dynamic visual stimuli. The underlying thesis is that goal directed targeting of attention may not be entirely voluntary, i.e. attentional targeting may be influenced by joint structural properties associated with dynamic stimuli, and thus, attention may be involuntarily controlled, at least in part by the design of the display dynamics. This dynamic attending approach is a biological view on how we allocate our attentional resources to dynamic information in the environment. There are three basic assumptions in this dynamic approach to understanding how humans interact with complex systems: (1) Attention is controlled, in part, by the combined spatial-timing structure of dynamic visual information, (2) Attentional energies are stimulated and "synchronized" with certain invariantly occurring dynamic space-time structures, and (3) The focus of the attention may speed up or slow down when there are abrupt accelerations or decelerations. It has been clearly demonstrated that certain spatial-time patterns are productive for the transfer of information from the display to the operator and other patterns can be very counterproductive for this task.

2. OBJECTIVE

The objective in this investigation is use the experimental platform involving dynamic attending with the approximate entropy measure on performance data involving a discrete sequence of tasks. A causality will be shown between the ApEn metric and the information compatibility (dynamic attending) of the presentation of data to the operator. The ApEn metric can be calculated in real time when the dynamic aspects of the display may change. Both the operator and experimenter are blind to the active display characteristics and how it evolves with time.

3. HYPOTHESIS

This study will show the effectiveness of the approximate entropy metric to identify when the display dynamics may change. The goal is to reject the null hypothesis H_0 at an α level of 0.05 such that:

H_0: There is no change in the approximate entropy measure during extreme high and low periods of information incompatibility (dynamic attending).

4. METHODS

Data from eight USAF subjects/contractors were collected as they performed a time estimation task. Two levels of information compatibility (dynamic attending) were presented to the subjects. These

difficulty levels were kept blind from both the experimenter and the subjects. Figure 1 illustrates the visual scene in which the subjects make estimates of the time perception.
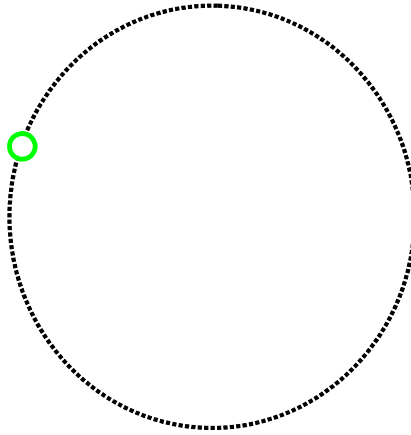

Figure 1 – Time Estimation Task

In Figure 1, the subject sees a pattern of stimuli being presented (small circle) in a certain timing pattern. The small circle traverses the larger circular pattern (the large circle is actually hidden) until the small circle changes color. After this color change, the subject has to estimate the time for the next appearance of the circle as being either "early", "late" or "on time" in comparison to the prior timing pattern. The subject responds with a mouse cursor as to one of the three choices (early, late or on time). From the works of Skelly et al., Table 1 displays the four possible paradigms that are known to produce different levels of dynamic attention or discordancy. Going down the table is in a direction of increasing task difficulty as determined from prior studies.

Figure 2 portrays an hypothesized trial from one possible set of runs. The task may be easy for a certain period of time and then, without notice to the subject or experimenter, the tasks may enter a region of increased difficulty for some period of time. Finally the task set may decrease its level of difficulty. The opposite sequence may also occur. The presentations of the task difficulty levels are counter balanced in the data presented here to eliminate confounding due to fatigue or ordering effects. Also in Figure 2 is a presumed plot of the variable, approximate entropy, drawn on the same time axis. The conjecture is that as the

Table 1 – Four Types of Dynamic Attending

Task Condition	Difficulty Level
Equal Time Equal Space	Easiest Task- Level 1
Equal Time	Next Most Difficult

23

Unequal Space	Task – Level 2
Unequal Time- Unequal Space	Third Most Difficult Task – Level 3
Unequal Time – Equal Space	Most Difficult Task – Level 4

difficulty of the task increases, the entropy would increase, accordingly. The entropy is evaluated by the degree of error made by the subjects as they respond to the question of whether the last appearance of the circle was either early, late, or on time. There are five levels of error (zero error, $\pm$ one unit and $\pm$ two units of error) that a subject can make. The approximate entropy depends on the absolute value of the level of the error made by the subject.
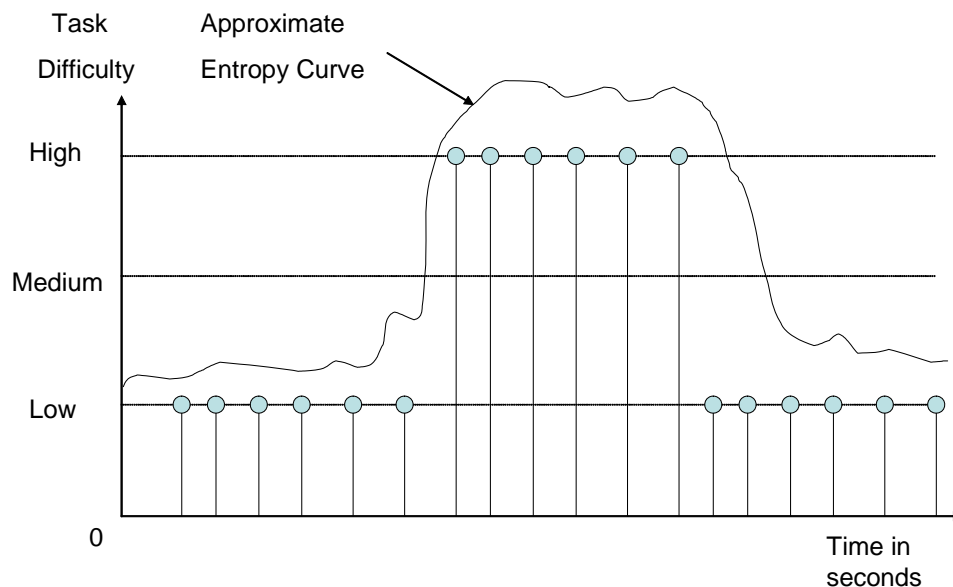
Figure 2 – An Example of The Temporal Presentation of Tasks

5. APPARATUS

A PC-based Borlan C program operated in DOS was employed to generate the experimental scenario and collect data. Data analysis was accomplished with MATLAB ™ with statistical analysis conducted with JMP 4.0 (SAS Institute). Figure 3 shows the facilities at the Air Force Research Laboratory to study dynamic attending issues. Figure 4 portrays one of the subjects in an early pretraining run.

Figure 3 – Experimental Setups to Perform Time Estimation Testing

6. EXPERIMENTAL DESIGN

Both genders of healthy USAF subjects/contractors participated in this experiment. This research had the goal of correlating the approximate entropy measure with the task difficulty of the underlying dynamical experimental conditions the operators were being exposed to. One data run of the training, the easy or hard task consisted of 64 trials. Frequent breaks

Figure 4 – Subject Running During the Time Estimation Task

were given to the subjects so that they spent no more than 60 minutes performing the testing each day. The dependent measure is the error (between the real world and the perceived value) in correctly classifying the differences in the time estimate when the last circle should appear. The approximate entropy measure was derived from this error signal in real time as described in the appendix. Errors are assigned as one unit if the subject guessed "On Time" when the actual stimulus was either early or late. An error of two units could be recorded if the subject responded "Late" when the actual stimulus

25

was early. The reverse situation would also produce two units of error, if it occurred. Other cases of 1 error unit were also possible.

7. RESULTS

We report data from eight subjects who have completed at least one training day and two additional data days of the time estimation testing. The analysis of the data is conducted as a within-subjects, full factorial design. There are two analyses to describe. The first analysis deals with the empirical performance results based on the errors accumulated. The second analysis deals with the use of the approximate entropy metric to distinguish the subject's response during the presentation of three levels of task difficulty (dynamic attending).

7.1   PERFORMANCE RESULTS

Similar to the work reported earlier by Skelly and colleagues (Skelly (2003), Skelly et al. (2000) and Jones and Skelly (1995)), we describe some basic tests on the data collected in this study. Figure 5 shows performance data from one of the subjects during three independent testing regimes. The absolute value of the error signal is plotted on the y axis, the trial number is on the x axis.
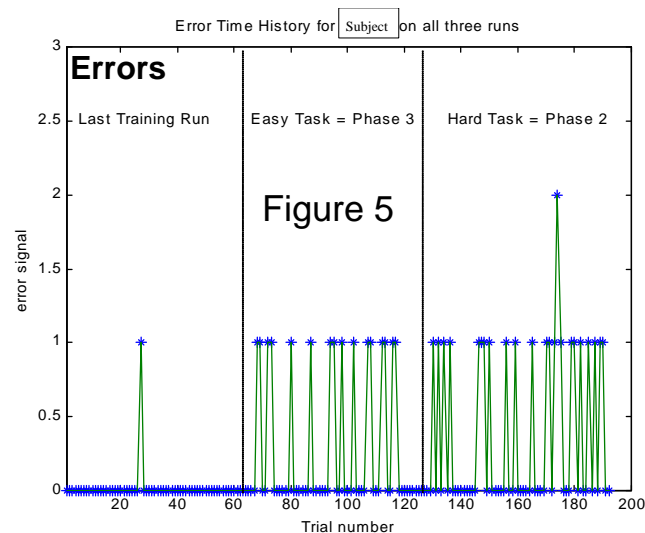
In Figure 5, for the first 64 trials (training run), the subject made one error of time judgment during this event. This was the final training run and subjects were required to have over 95% correct time judgments prior to entering into the latter phases of testing. During the training and data runs, subjects had 5 brief breaks. The total duration of a run was 19 minutes to complete all 64 trials. After completion of the training run, the subjects would either perform an "easy" task for 64 trials or a "hard" task. The distinction between easy or hard task was defined by Dr. Skelly (different techniques of dynamic attending) and was randomly presented to the subjects in blocks of 64 trials. In Figure 5, the easier task was presented prior to the more difficult task. In Figure 5, the number of errors was found to be:

26

Training Task: 1 error in 64 trials.

Easy Task: 9 errors in 64 trials.

Hard Task: 18 errors in 64 trials.

Of course, if a subject received the treatments easy and then hard on one data day, then he must also receive the reverse order of hard and easy on the next data day to counterbalance the experimental conditions. It should be emphasized that the task complexity (bits) and the bit rate of presentation of information (Fitts' law or bits/second) were identical for the easy and hard tasks. It was only the dynamic presentation of the information that was different between the tasks. Since brevity must be the style here, the first analysis deals with performance in terms of the number of errors that occurred. This is determined versus the levels of task difficulty exhibited in Table 2. The mean and SE of the number of errors across all 8 subjects accumulated during the 64 trials is displayed. A one-way ANOVA on the difference of means of the dependent measure of errors in time perception across the task difficulty levels is provided to examine the efficacy of the experimental scenario to elicit a performance change.

Table 2 – Mean and SE of Errors Observed

Training		Easy Task		Hard Task	
Mean	SE	Mean	SE	Mean	SE
1.91	1.35	9.64	1.35	20.54	1.35

Table 3 illustrates the results of a one-way ANOVA (t-test of means – all pairs Tukey-Kramer) based on JMP 4.0 analysis (SAS Institute, 2004).

Table 3 – One Way ANOVA for Errors

Source	DoF	Σ Squares	Mean Squares	F-Ratio	Prob>F
Task Level	2	1928.8	964.394	48.045	<.0001
Error	30	602.2	20.073		
C. Total	32	2530.96			

  Figure 6 displays the output of the comparison of means of errors across all 8 subjects which clearly (with the one-way ANOVA results of Table 2) demonstrates the efficacy of the experimental paradigm to elicit a performance change as predicted by the prior works of Skelly et al.
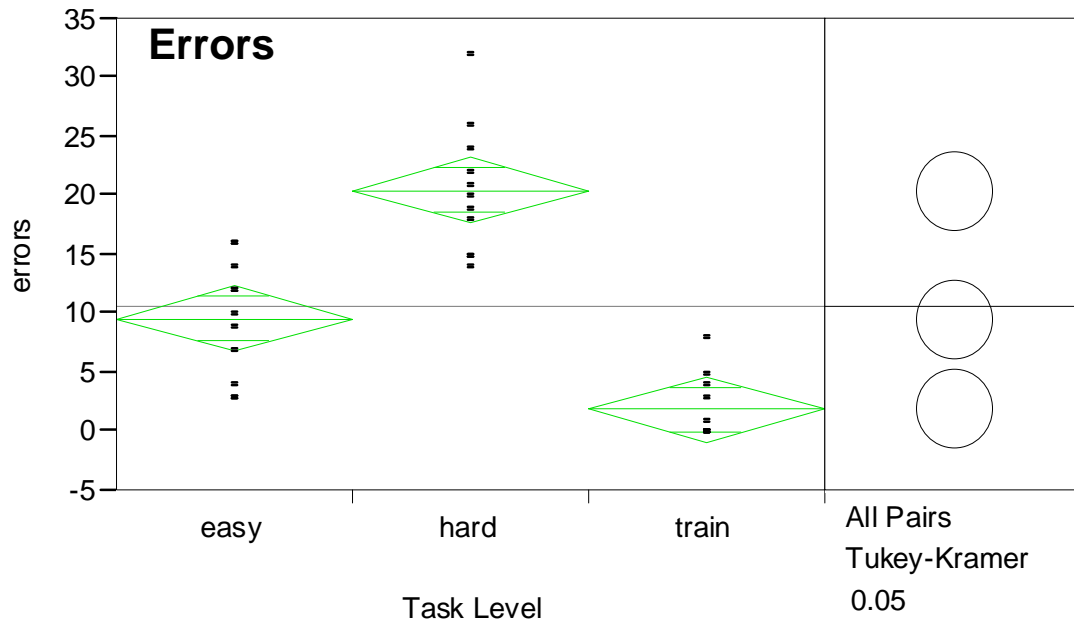
Figure 6 – Distribution of Errors for the Three Levels of Task Difficulty as Defined by Skelly et al.

7.2   RESULTS FROM ApEn ANALYSIS

A second goal of this investigation was to show the efficacy of the ApEn metric to demonstrate that the disorder in the operator's response is also affected in a dynamic time sense. The platform for the experimental design (dynamic attending) discussed in section 7.1 clearly provides an investigational scenario to manipulate the disorder in the response of the operator in terms of task difficulty. A plot of the ApEn versus time is shown in Figure 7 for one subject and the three runs: training, the easy task, and the hard task. The Appendix describes the implementation issues for the generation of the ApEn calculation portrayed in Figure 7.   Other technical details on the ApEn for the use of determination of irregularity in real time empirical data can also be found in Repperger, et al., 2004, published in this same conference proceedings. In Figure 7 it is noted that three different levels of ApEn seem to appear at each level of task difficulty. Table 4 shows the results of averages of the ApEn metric over the 8 subjects. The averages across subjects were conducted for the mean ApEn value during the 64 trials within the same task difficulty interval as indicated in Figure 7.

28
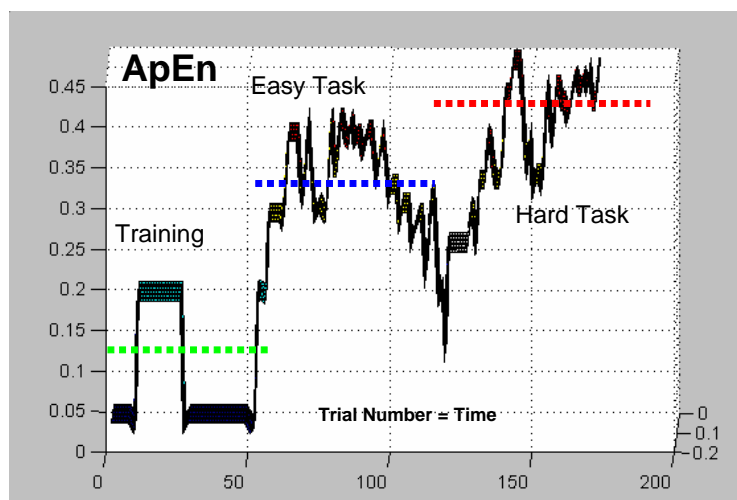
Figure 7 – Side view of ApEn Plot versus trial number for one subject.

Table 4 – Mean and SE of ApEn Calculated

Training		Easy Task		Hard Task	
Mean	SE	Mean	SE	Mean	SE
0.074	0.02	0.182	0.02	0.247	0.02

Figure 8 illustrates a three dimensional plot of the ApEn function versus time for the same data as displayed in Figure 7 with the third axis representing r = the threshold window as discussed in the appendix.

Figure 8 – ApEn Plot versus Trial Number and r = Threshold Window

Table 5 illustrates the results of the one-way ANOVA (t-test between means – Tukey-Kramer) based on JMP 4.0 analysis (SAS Institute, 2002) for the ApEn metric in Figures 7-8.

29

Table 5 – One-Way ANOVA for Magnitude of ApEn
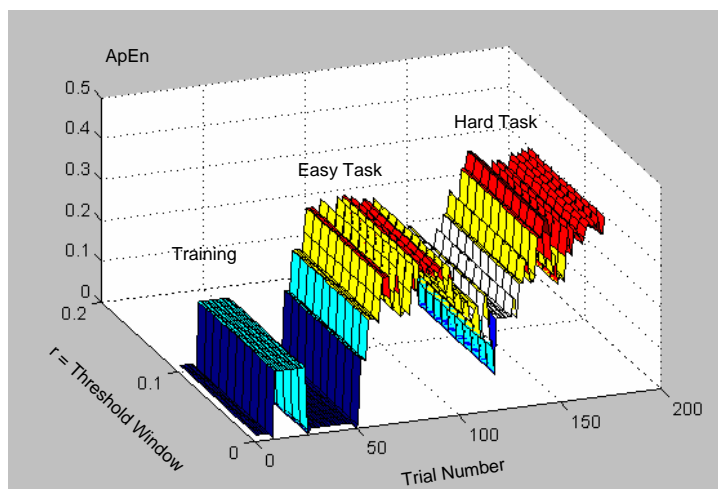
Source	DoF	Σ Squares	Mean Squares	F-Ratio	Prob>F
Task Level	2	0.3031	0.1515	21.16	<.0001
Error	30	0.2148	0.0072		
C. Total	32	0.5179			

Figure 9 portrays the JMP 4.0 analysis output as discussed in Table 5 for a one-way ANOVA on the dependent measure of magnitude of ApEn during each of the three levels of task difficulty considered in this experimental study.

8. DISCUSSION

Tables 3 and 5, presented in section 7, clearly show that the efficacy of the experimental paradigm to elicit a performance change as well as the ability of the ApEn metric to provide sufficient sensitivity to sense this change. Clearly all the levels of different dynamic attending
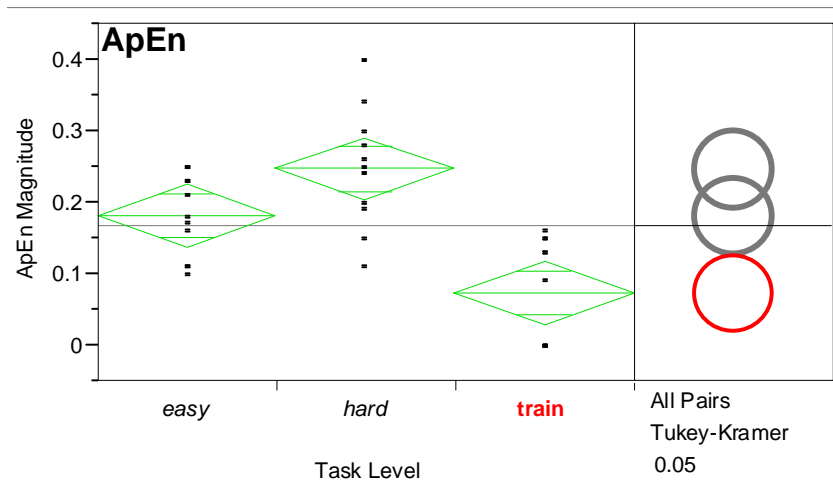
Figure 9 – Distribution of Means of ApEn values over levels of Task Difficulty Across 8 Subjects.

difficulty can be distinguished either through the performance changes observed or via the ApEn metric which can be developed online in real time.

In summary, a preliminary study has been conducted in time estimation when the dynamic display has a known change in dynamic attending capability, but with identical task complexity and complexity rate. The efficacy of the approximate entropy measure to capture and characterize this dynamic change has been evaluated. This important real time measure (ApEn) provides a valuable means of describing how humans deal with complex systems which may have characteristics that vary with time.

30

9. CONCLUSIONS

Studying dynamic displays is a difficult task. This preliminary study looked at a measure of this interaction in terms of the disorder of the response of the subjects (entropy measure). Additional studies will examine display conditions and the ability of this measure or other metrics to capture the true dynamic interaction of the operator within a changing dynamic environment.

10. REFERENCES to [4]

Fitts, P. M. (1954). "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.*, **67**: pp. 103- 112.

Gevers, E., Pincus, S. M., Iaian, C. A., Robinson, A. F., and Veldhuis, J. D. (1998). "Differential Orderliness of the GH release Process in Castrate Male and Female Rats," *The American Journal of Physiology*, pp. R437-R444.

Jones, M. R. and Skelly, J. J. (1993). "The Role of Event Time in Attending," *Time and Society*, Sage, **2** (1): pp. 107-128.

Pincus, S. M., and Viscarello, R.R. (1992). "Approximate Entropy: A Regularity Measure for Fetal Heart rate Analysis," *Obstetrics and Gynecology*, **79**, (2), February pp. 249-255.

Pincus, S. M. (2003). "Quantitative Assessment Strategies and Issues for Mood and Other Psychiatric Serial Study Data," *Bipolar Disorders*, **5**, pp. 287-295.

Repperger, D. W., Albery, W. B, and Tripp, L. D. (2004). "Approximate Entropy as an Assessment Tool for System Complexity and Performance Valuation in Human-Machine Systems," *9th IFAC/IFIP /IFORS/IEA Symposium, Analysis, Design, and Evaluation of Human-Machine Systems*, Atlanta, Georgia.

Skelly, J. J. (2003). "Time: Our Lost Dimension of Interface Design," *Proceedings of the BRIMS 2003 Conference*, Scottsdale, AZ, May 11-15.

Skelly, J. J., Jones, M. R. Goodyear, C. D., Rose, M. M. (2000). "Attentional Pacing and Temporal Capture in Slow Visual Sequences," AFRL-HE- WP-TR-2003-0078, September.

APPENDIX of [4] - DETAILS REGARDING IMPLEMENTAION OF THE ApEn METRIC

Following the other reference (Repperger, et al., 2004, published in [3]), the choice was made of m=1, n=10, and the window r was adjusted to be approximately 25% of the standard deviation of the data, as suggested by Pincus and Viscarello (1992). The difficulty in implementing the ApEn measure is because this is a discrete task, rather than a continuous time series sampled at a uniform sampling rate. To convert this time estimation experimental paradigm into a procedure amenable to analysis using the ApEn metric, one needs to look at the error signal, e.g. as it appears in Figure 5. The kind suggestion[2] was made that to convert a sequence of discrete tasks into a means amenable for calculation by the ApEn metric, it is only necessary to concatenate the error trials versus time, as shown in Figure 5. The ApEn analysis considers the sequence of 64 x 3 = 192 trials as a single time series. The independent variable is the trial number (1-192). The dependent measure is the absolute value of the error signal (which is 0, 1, or 2 units). The ApEn analysis was applied to the data by treating the sequence of 192 trials as 192 uniform samples of time data. Since the ApEn analysis is independent of the sampling time, it does not impact the analysis procedure used here.

On the Absolute Orientation Problem in Computer Vision

Rodney G. Roberts
Dept. of Electrical and Computer Eng.
Florida State University
Tallahassee, FL, USA
rroberts@eng.fsu.edu

Daniel W. Repperger
Air Force Research Laboratory
711 HPW AFRL/RHCV
Wright-Patterson AFB, OH 45433
Daniel.Repperger@wpafb.af.mil

Abstract—An important problem in computer vision is to determine the orientation of a rigid body in an image. This can be accomplished by matching points or line segments that naturally appear on the object. Several elegant and computationally fast algorithms based on the singular value decomposition and quaternions have been introduced to solve this problem. In this article, the authors first examine the important special case of identifying the attitude of 2D objects and introduce a particularly elegant solution based on the mathematical structure of the complex plane. Motivated by this simple solution to the 2D case, a new derivation of the 3D case based on the polar decomposition is presented. This derivation is in many ways more natural than previous derivations, particularly when the model and data contain no noise.

Index Terms—Absolute orientation, least squares, polar decomposition.

I. INTRODUCTION

A fundamental problem in computer vision is the determination of the orientation of a rigid object. An effective approach to this problem is to match a set of points on the object with the corresponding points on a model. In particular, the following mathematical problem appears in a number of references [1]-[6].

Two point sets $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_i\}$ of N vectors in the plane or in 3-space are related by

$$\mathbf{b}_i = R\mathbf{a}_i + \mathbf{t} + \mathbf{n}_i \tag{1}$$

where R is a rotation matrix, $\mathbf{t}$ is a translation vector, and $\mathbf{n}_i$ is a noise term. The set $\{\mathbf{a}_i\}$ corresponds to the location of several specified points on a model of the object while $\{\mathbf{b}_i\}$ represents the corresponding points on the object in an image. The goal is to determine R and $\mathbf{t}$ to minimize

$$F(\mathbf{t}, R) = \sum_{i=1}^{N} \|\mathbf{b}_i - (R\mathbf{a}_i + \mathbf{t})\|^2 \tag{2}$$

where $\|\cdot\|$ is the standard 2-norm. The vector $\mathbf{t}$ and rotation matrix R represent the location and orientation of the object in the image. The problem of matching line segments, although more complicated, results in essentially the same type of optimization problem [2].

Several approaches to this problem have been described in the literature including matrix-based solutions such as the singular value decomposition (SVD) [1]-[3] and quaternion-based solutions [4],[5]. In the next section, we introduce a new proof for the 2D case based on simple properties of complex numbers. In particular, it is shown that the solution for the orientation of the object is given by the polar form of a particular complex number. In Section III, this solution motivates a natural solution to the general 3D problem based on the polar decomposition of a particular matrix. In fact, the polar decomposition is the obvious solution when no noise is present in the problem. After solving for the case when there is noise in the data, we provide a new proof that the same solution holds when there is noise in both the model and the data. Lastly, conclusions appear in Section IV.

II. SOLVING THE PLANAR CASE USING THE COMPLEX PLANE

A popular solution to the orientation problem is based on the singular value decomposition (SVD). However, as one would expect, the much simpler 2D case does not require the sophistication of an SVD, not only because of the smaller dimension size, but more importantly, because of the commutativity of the rotation operation. In this case, it is convenient to formulate the problem in terms of complex numbers. Suppose that $\mathbf{a} = \begin{bmatrix} a_x & a_y \end{bmatrix}^T$ and $\mathbf{b} = \begin{bmatrix} b_x & b_y \end{bmatrix}^T$ are vectors in the plane. If we write these vectors in complex number notation as $a = a_x + ja_y$ and $b = b_x + jb_y$, then the inner product $\mathbf{a} \cdot \mathbf{b}$ of the two vectors is given in complex number notation as $\mathrm{Re}(a^*b)$ where a^* denotes the complex conjugate of the complex number a and where $\mathrm{Re}(z)$ denotes the real part of z. Furthermore, the norm squared $\|\mathbf{a}\|^2$ of the vector $\mathbf{a}$ is given by $|a|^2 = a^*a$, and the rotation matrix corresponding to a counterclockwise rotation of θ radians is given by the complex number $e^{j\theta}$. Based on this formulation, the goal is to minimize the objective function

$$F(t, e^{j\theta}) = \frac{1}{N} \sum_{i=1}^{N} |b_i - (e^{j\theta}a_i + t)|^2 \tag{3}$$

where the complex number t and the real number $\theta \in [0, 2\pi)$ represent the position and orientation of the object, respectively. Since determining the optimal $t_0(\theta)$ for a given θ is a routine least squares calculation, we merely state the result

that

$$t_0(\theta) = \bar{b} - e^{j\theta}\bar{a} \tag{4}$$

where $\bar{b} = \frac{1}{N}\sum_{i=1}^{N} b_i$ and $\bar{a} = \frac{1}{N}\sum_{i=1}^{N} a_i$. We define $\tilde{a}_i = a_i - \bar{a}$ and $\tilde{b}_i = b_i - \bar{b}$ and say that $\{\tilde{a}_i\}$ and $\{\tilde{b}_i\}$ are the unbiased versions of $\{a_i\}$ and $\{b_i\}$, respectively. It then follows that

$$\begin{aligned} F(t, e^{j\theta}) &= \frac{1}{N}\sum_{i=1}^{N} |\tilde{b}_i - e^{j\theta}\tilde{a}_i|^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \left[|\tilde{b}_i|^2 + |\tilde{a}_i|^2 - 2\mathrm{Re}((\tilde{b}_i)^* e^{j\theta}\tilde{a}_i) \right]. \end{aligned} \tag{5}$$

Since $\frac{1}{N}\sum_{i=1}^{N} \left[|\tilde{b}_i|^2 + |\tilde{a}_i|^2 \right]$ is fixed by the data, we want to maximize $\mathrm{Re}\left[\left(\frac{1}{N}\sum_{i=1}^{N} (\tilde{b}_i)^* \tilde{a}_i \right) e^{j\theta} \right]$ with respect to θ, which is clearly achieved by choosing

$$\theta = -\arg\left(\sum_{i=1}^{N} (\tilde{b}_i)^* \tilde{a}_i \right) = \arg\left(\sum_{i=1}^{N} \tilde{a}_i^* \tilde{b}_i \right) \tag{6}$$

where $\arg(z)$ denotes the argument of the complex number z and has range $[0, 2\pi)$. The location and orientation of the object is then given by (4) and (6), respectively.

Another approach, that will serve as a guide in the next section to solve the 3D case, is to write

$$\begin{aligned} F(t, e^{j\theta}) &= \frac{1}{N}\sum_{i=1}^{N} \left[|\tilde{b}_i|^2 + |\tilde{a}_i|^2 \right] - 2\mathrm{Re}(c^* e^{j\theta}) \\ &= \frac{1}{N}\sum_{i=1}^{N} \left[|\tilde{b}_i|^2 + |\tilde{a}_i|^2 \right] - |c|^2 - 1 + |c - e^{j\theta}|^2 \end{aligned} \tag{7}$$

where $c = \frac{1}{N}\sum_{i=1}^{N} \tilde{a}_i^* \tilde{b}_i$. We thus want to minimize $|c - e^{j\theta}|^2$, which is clearly achieved by $\theta = \arg(c)$ or, equivalently, $\theta = \arg\left(\sum_{i=1}^{N} \tilde{a}_i^* \tilde{b}_i \right)$.

III. SOLVING THE GENERAL CASE USING THE POLAR DECOMPOSITION

A popular approach to solving the 3D case is based on quaternions [4], [5]. While quaternions are a generalization of complex numbers, the complex number approach of the previous section more naturally leads to a matrix solution based on the polar decomposition. We begin by reformulating the problem in matrix notation by letting $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_N \end{bmatrix}$, $B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_N \end{bmatrix}$, and $N = \begin{bmatrix} \mathbf{n}_1 & \cdots & \mathbf{n}_N \end{bmatrix}$. Equation (1) then becomes

$$B = RA + \mathbf{t}\mathbf{e}^T + N \tag{8}$$

where $\mathbf{e} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T$. The optimization problem then becomes to minimize

$$F(\mathbf{t}, R) = \|B - (RA + \mathbf{t}\mathbf{e}^T)\|_F^2 \tag{9}$$

subject to R being a rotation matrix where $\|\cdot\|_F$ denotes the Frobenius norm, which is given by the square root of the sum of the squares of the matrix elements.

A. The Noise-free Case

We first examine the simplest possible case, i.e., when no noise is present. In this ideal case, we have an exact equality, which can be written in matrix form as

$$RA + \mathbf{t}\mathbf{e}^T = B \tag{10}$$

where $\mathbf{e} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T$. The translation term $\mathbf{t}$ can be found by post-multiplying (10) by $\frac{1}{N}\mathbf{e}$ to obtain

$$\mathbf{t} = \bar{\mathbf{b}} - R\bar{\mathbf{a}} \tag{11}$$

where $\bar{\mathbf{a}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{a}_i$ and $\bar{\mathbf{b}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{b}_i$. Writing $\tilde{A} = A - \bar{\mathbf{a}}\mathbf{e}^T$ and $\tilde{B} = B - \bar{\mathbf{b}}\mathbf{e}^T$, we have $R\tilde{A} = \tilde{B}$ or, equivalently, $\tilde{A}^T R^T = \tilde{B}^T$. Lastly, pre-multiplying both sides by $\tilde{A}$ gives

$$\tilde{A}\tilde{A}^T R^T = \tilde{A}\tilde{B}^T. \tag{12}$$

We assume that the 3×3 matrix $\tilde{A}\tilde{B}^T$ has full rank. Then the unique polar decomposition of $\tilde{A}\tilde{B}^T = PQ$ is given by the left hand side of (12) where $P = \tilde{A}\tilde{A}^T$ is positive definite and $Q = R^T$ is an orthogonal matrix. Since $P = \tilde{A}\tilde{A}^T$ is positive definite, it follows that the determinants of $\tilde{A}\tilde{B}^T$ and Q have the same sign so that $R = Q^T$ is a rotation matrix if and only if $\det(\tilde{A}\tilde{B}^T) > 0$. If $\det(\tilde{A}\tilde{B}^T) < 0$, then R is a reflection matrix and the orientation problem is ill-defined.

The polar decomposition is a natural solution to the problem when no noise is present. Before continuing to the more general case, we observe that the solution is particularly simple when $\tilde{A}$ has the property that $\tilde{A}\tilde{A}^T = kI$. In that case, we merely scale $\tilde{A}\tilde{B}^T$ to obtain an orthogonal matrix. An example of this occurs when the columns of A correspond to the vertices of a platonic solid. Unfortunately, choosing a data matrix so that $\tilde{A}$ has this property may result in a numerically unstable solution when noise is present.

B. Noise in Only the Data

We now return to the problem formulated in (9), which can be rewritten as

$$F(\mathbf{t}, R) = \|\mathbf{t}\mathbf{e}^T - (B - RA)\|_F^2. \tag{13}$$

The optimal solution for $\mathbf{t}$ for a fixed rotation matrix R is given by the pseudoinverse solution

$$\mathbf{t} = (B - RA)(\mathbf{e}^T)^+ = \frac{1}{N}(B - RA)\mathbf{e} \tag{14}$$

so that it follows that $\mathbf{t}$ is once again given by (11). Substituting this expression into (13) gives

$$\begin{aligned} \|R\tilde{A} - \tilde{B}\|_F^2 &= \|\tilde{A}\|_F^2 - 2\mathrm{tr}(\tilde{B}^T R\tilde{A}) + \|\tilde{B}\|_F^2 \\ &= \|\tilde{A}\|_F^2 + \|\tilde{B}\|_F^2 - 2\mathrm{tr}(\tilde{A}\tilde{B}^T R). \end{aligned} \tag{15}$$

Applying the slip-in/slip-out method, we can write our objective function as

$$\begin{aligned} \|R\tilde{A} - \tilde{B}\|_F^2 &= \|\tilde{A}\|_F^2 + \|\tilde{B}\|_F^2 - \|\tilde{A}\tilde{B}^T\|_F^2 - \|R\|_F^2 \\ &\quad + \|\tilde{A}\tilde{B}^T\|_F^2 - 2\mathrm{tr}(\tilde{A}\tilde{B}^T R) + \|R\|_F^2 \\ &= \|\tilde{A}\|_F^2 + \|\tilde{B}\|_F^2 - \|\tilde{A}\tilde{B}^T\|_F^2 - n \\ &\quad + \|\tilde{A}\tilde{B}^T - R\|_F^2. \end{aligned} \tag{16}$$

Note the similarity of (16) with (7). The first four terms in the last equality of (16) are independent of R, so our problem becomes the optimization of $\|\tilde{A}\tilde{B}^T - R\|_F^2$.

Note the special structure of this final version of the optimization problem. We seek an orthogonal matrix that is closest to the 3×3 matrix $\tilde{A}\tilde{B}^T$. More generally, we consider the problem of determining

$$U_0 = \arg \min_{U \in O(n)} \|M - U\|_F \qquad (17)$$

where M is an $n \times n$ full rank matrix. We solve this problem by first examining two special cases. First, suppose M is a diagonal matrix: $M = \text{diag}(d_1, \ldots, d_n)$. Then

$$\begin{aligned} \|M - U\|_F^2 &= \|D - U\|_F^2 \\ &= \|D\|_F^2 + \|U\|_F^2 - 2\text{tr}(DU) \\ &= \sum_{i=1}^n d_i^2 + n - 2\sum_{i=1}^n d_i u_{ii}. \end{aligned} \qquad (18)$$

Since $\sum_{i=1}^n d_i^2$ and n are fixed, we need to maximize $\sum_{i=1}^n d_i u_{ii}$. Since $-1 \leq u_{ii} \leq 1$, we have that $\sum_{i=1}^n d_i u_{ii} \leq \sum_{i=1}^n |d_i|$ and it follows that $U_0 = \text{diag}(\text{sgn}(d_1), \ldots, \text{sgn}(d_n)) \in O(n)$. Furthermore, if D is positive definite, i.e., if $d_i > 0$ for $i = 1, \ldots, n$, then $U_0 = I$.

For our second case, suppose that M is symmetric. Then we can write M as $M = VDV^T$ where $V \in O(n)$ and $D = \text{diag}(d_1, \ldots, d_n)$. We then have

$$\|M - U\|_F = \|VDV^T - U\|_F = \|D - VUV^T\|_F \qquad (19)$$

where the second equality in (19) follows from the fact that the Frobenius norm is invariant under pre- and post-multiplication by orthogonal matrices. By the first case, it follows that the optimal $U_0 \in O(n)$ is given by $V^T U_0 V = \text{diag}(\text{sgn}(d_1), \ldots, \text{sgn}(d_n))$, i.e., the optimal orthogonal matrix is given by $U_0 = V\text{diag}(\text{sgn}(d_1), \ldots, \text{sgn}(d_n))V^T$. For the important case when M is symmetric positive definite, this becomes $U_0 = I$.

We are now ready for the general case when M is an arbitrary full rank $n \times n$ matrix. In this case, we can write M in its polar form: $M = PQ$ where P is symmetric, positive definite and $Q \in O(n)$. Then $\|M - U\|_F = \|PQ - U\|_F = \|P - UQ^T\|_F$, which is minimized over $U \in O(n)$ by the orthogonal matrix U_0 where $U_0 Q^T = I$, i.e., $U_0 = Q$.

We thus conclude that the optimal solution for determining the orientation of the object in question is given by the orthogonal matrix $R = Q^T$ where PQ is the unique polar decomposition of $\tilde{A}\tilde{B}^T$. We note that when the polar decomposition has been mentioned in the literature as a solution, it has primarily appeared as an afterthought of the SVD solution. While these solutions are arguably equivalent, the SVD solution does not appear to be as natural as the polar decomposition solution. It is important to note once again that R is a rotation matrix if and only if $\det(Q) = 1$; otherwise, R is a reflection matrix.

C. Noise in Both the Model and Data

We next examine the case when there is noise in both the model and the data. In particular, we consider the following problem described in [6], given here with slightly different notation. Suppose we have two sets of N noisy observations given by two $3 \times N$ matrices A and B. We assume that the correspondence problem has already been solved so that the corresponding points are in the same order in A and B and that the translation of the object has already been determined. Our problem then is to find a rotation matrix R and perturbations δA and δB which satisfy

$$R(A + \delta A) = B + \delta B \qquad (20)$$

such that $\|\delta A\|_F^2 + \|\delta B\|_F^2$ is minimized. The perturbations δA and δB correspond to noise in the model and data, respectively. This problem, presented in slightly different notation, was solved by Goryn and Hein in [6], but the solution presented there relies heavily on the introduction of some non-obvious substitutions. We present a more natural and intuitive derivation. We begin by first noting that unlike δB, the term δA appears in (20) as $R\delta A$, suggesting that it may be better to rewrite the cost function as $\|\delta A\|_F^2 + \|\delta B\|_F^2 = \|R\delta A\|_F^2 + \|\delta B\|_F^2$, where equality follows from the fact that R is orthogonal. Next, equation (20) can be written as $R\delta A - \delta B = -(RA - B)$. We thus want to minimize $\|R\delta A\|_F^2 + \|\delta B\|_F^2$ subject to $R\delta A - \delta B = -(RA - B)$.

The scalar version of the preceding problem statement suggests a solution to this constrained optimization problem. In the scalar case, we want to minimize $x^2 + y^2$ over the scalars x, y, z subject to the constraint $ax + by = f(z)$ where $f(z) = -cz + d$ and where a, b, c, d are fixed parameters. Geometrically, for a fixed z, the constraint can be interpreted as the equation of a line and the optimal solution over x, y would then correspond to the point on that line which is closest to the origin. We then want to choose z to place the line as close to the origin as possible, which is achieved by minimizing $|f(z)| = |cz - d|$. Once this is done, the optimal x and y can be determined. This suggests minimizing the norm of $RA - B$ in the non-scalar case.

We now present a formal derivation of the solution for the non-scalar case. The constraint (20) in the general case can be written in matrix notation as

$$\begin{bmatrix} \frac{1}{\sqrt{2}}I & \frac{-1}{\sqrt{2}}I \end{bmatrix} \begin{bmatrix} R\delta A \\ \delta B \end{bmatrix} = \frac{-1}{\sqrt{2}}(RA - B) \qquad (21)$$

where the $1/\sqrt{2}$ term is included so that the rows of the matrix on the left are not only mutually orthogonal, but are also normalized. This suggests augmenting the matrix so that it becomes orthogonal:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} I & -I \\ I & I \end{bmatrix} \begin{bmatrix} R\delta A \\ \delta B \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -(RA - B) \\ R\delta A + \delta B \end{bmatrix}. \qquad (22)$$

Since the Frobenius norm is invariant under multiplication by orthogonal matrices, we have that the cost function $\|\delta A\|_F^2 + \|\delta B\|_F^2$ is given by

$$\|R\delta A\|_F^2 + \|\delta B\|_F^2 = \frac{1}{2}\|RA - B\|_F^2 + \frac{1}{2}\|R\delta A + \delta B\|_F^2, \quad (23)$$

which is clearly minimized by setting $\delta B = -R\delta A$ and minimizing $\|RA-B\|_F^2$ over the family of orthogonal matrices R so that R is determined in the same manner as before.

IV. CONCLUSION

In this article, we have presented the polar decomposition as the natural method for solving the absolute orientation problem. Previously, the polar decomposition approach was mentioned in the literature as merely an afterthought of the SVD approach. The polar decomposition method for the 3D case was motivated by a complex number approach to the 2D case, which is interesting in its own right. Lastly, we have provided a simpler and more natural proof that the same solution also holds for a noisy model.

ACKNOWLEDGMENT

REFERENCES

[1] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. PAMI*, vol. 9, no. 5, pp. 698–700, Sept. 1987.

[2] B. Kamgar-Paris and B. Kamgar-Paris, "Matching sets of 3D line segments with application to polygonal arc matching," *IEEE Trans. PAMI*, vol. 19, no. 10, pp. 1090–1099, Oct. 1997.

[3] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America A*, vol. 5, no. 7, pp. 1127-1638, July 1988.

[4] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, 4(4):629-642, April 1987.

[5] M. D. Wheeler and K. Ikeuchi, "Iterative estimation of rotation and translation using the quaternion," Technical Report CMU-CS-95-215, Carnegie-Mellon University, December 1995.

[6] D. Goryn and S. Hein, "On the estimation of rigid body rotation from noisy data, " *IEEE Trans. PAMI*, vol. 17, no. 12, pp. 1219–1220, Dec. 1995.